“You Should Not or You Must Not”

The Impact of Goal-Type on the

Infliction of Harm on Members of Deviant Groups

Dissertation

Zur Erlangung des akademischen Grades

Doctor philosophiae (Dr. phil.)

Vorgelegt dem Rat der Fakultät

für Sozial- und Verhaltenswissenschaften

der Friedrich-Schiller-Universität Jena

Von Dipl.-Sozialwirt Bastian Lücke

geboren am 22.05.1979 in Hannover

Gutachter

1. _____

2. _____

Tag des Kolloquiums: _____

Content

Figures

Tables

1 Introduction

The present research is concerned with factors determining the extent of harm that is inflicted on members of deviant groups. One essential characteristic of a group is the pursuit of a common group goal. However, those who refrain from the pursuit of this group goal are likely to be perceived as deviants by their fellow group members. It is argued that the representation of a group goal either as an ideal, that should be aspired, or as a threshold, that has to be unconditionally reached, is a crucial determinant of the extent of harm that is inflicted on members of a deviant group. Two examples might illustrate this concern.

Example 1:

In a strike workers are united by the common goal of fighting for better working conditions and higher wages for all workers. But there are individual rationales not to join in a strike. Some workers may shy away from an open confrontation with their employer and continue to work. But those workers who do not join the strike will be considered strike-breakers by their fellow workers. Strike-breakers are often the target of verbal abuse, social exclusion, beatings and even premeditated murder because of their deviation from the common group goal of striking. Sometimes even uninvolved third parties such as the families of strike-breakers are threatened. But why is that in some cases simple name-calling is perceived to be the adequate reaction towards the deviant group of strike-breakers and sometimes cold-blooded murder?

Example 2:

It is one of the most fundamental group endeavors to protect one´s country against an enemy. However, many religious groups such as Franciscans, Mennonites, Quakers or Jehova´s Witnesses have been refusing to fight in armed conflicts because of their religious beliefs. Many kinds of harm have been inflicted on these religious groups in response. These forms of harm have been ranging from simple fees, to significantly prolonged alternative social service, social exclusion in the form of imprisonment, to extreme forms of punishment such as summary executions. But what determines if a group that is refusing to fight in a war is simply fined or killed for this deviation from a common group goal?

These two examples share a number of basic similarities. Both examples are based on an important group goal that requires the collective effort of its group members. This group

endeavor can be a strike or the defense of one´s country. But in both examples a number of group members deviate from this group goal. These deviants may constitute a group in advance as the religious groups in the second example or may be defined as a group by their coherent behavior of deviating from the group goal as the non-striking workers in the first example. The conceivable negative reactions toward members of those deviant groups range in both examples from relatively mild forms of punishment such as name calling to severe forms of harm such as physical violence and murder. It is suggested that the way the common group goal is represented has a crucial impact on the infliction of harm on those deviants.

The group goal of striking might be represented as an ideal, a *maximal goal*. This goal should be aspired as much as possible. The negative reactions towards a group of workers deviating from this group goal of striking by continuing their work are expected to be graded in this case: The more a group of workers deviates from the group goal of striking, the more harm is inflicted on them. But the group goal of striking could also be represented as a dichotomous *minimal goal*. A minimal goal is either fully achieved or not at all. It constitutes a normative border that must not be crossed. Picket lines might be interpreted as a behavioral manifestation of such a dichotomous, *minimal goal* representation of the group goal of striking. Any group deviating from the group goal of striking instead of working - crossing the picket line - will be confronted with the infliction of severe harm. The degree of their transgression - how far or how long the picket line is crossed - does not affect the extent of harm that is inflicted on them. The extent of harm will be severe, no matter how far or for how long they cross the picket line.

The same logic can be applied to the example of the common group goal of fighting for ones country. This group goal might be represented as an ideal *maximal goal*. The degree of harm that is inflicted on a group of deviants is relative, depending on the degree of their perceived deviation from this goal. The greater their perceived deviation from the group goal of fighting for one´s country, the more severe the extent of harm, that is inflicted on this group of deviants. However, the group goal of fighting for one´s country may also be represented as a minimal group goal, a goal that just has to be achieved - for example during times of war. In this case any deviation from this *minimal group goal* at all will be perceived as a severe transgression and will result in the infliction of severe forms of harm such as collective executions.

The same deviation from the very same group goal may lead to the infliction of different degrees of harm on a group of deviants, depending on the representation of the group goal as

either a minimal or a maximal goal. A group deviating from a group goal, that is represented as a *maximal goal* is expected to be punished relative to the degree of their perceived deviation. However, any deviation from a *minimal group goal* is perceived as a severe transgression - because of its dichotomous structure. Therefore, the very same group would be punished severely for any deviation from the group goal - regardless of the degree of their deviation - if this group goal was represented as a minimal goal. It is therefore the aim of the current research to scrutinize this expected impact of the representation of a group goal as either minimal or maximal on the infliction of harm on members of a group deviating from a group goal.

2 Theory

2.1 Deviant Groups

Both initial examples describe the infliction of harm on members of a deviant group. But what is the difference between a group that is perceived as merely different and a group that is perceived as deviant? First it has to be pointed out that the perception of deviance is not rooted in the deviants themselves or their behavior per se: "...social groups create deviance by making rules whose infraction creates deviance, and by applying those roles to particular people […] deviance is not a quality of the act the person commits, but rather a consequence of the application by other of rules and sanctions to an 'offender.'" (Becker, 1963). However, the question remains how a group defines which rules are perceived as binding for all of its members and what therefore determines when a group of individuals is perceived to merely *differ* or to *deviate* from these rules.

The Ingroup Projection Model (Mummendey & Wenzel, 1999) constitutes an attempt to explain the evaluation of Outgroups (OG) on the basis of Social Comparison Theory (e.g. Festinger, 1954) and Self Categorization Theory (Turner, Hogg, Oakes, Reicher, & Wetherell, 1987). The model assumes that groups are compared with reference to an inclusive superordinate group (SOG) and that members of an Ingroup (IG) tend to generalize IG-attributes to the prototype of this SOG (Ingroup-projection). The prototype of the SOG defines how a member of the SOG *ought* to be. Consequently the IG is perceived as more prototypical than the OG (higher perceived relative IG-prototypicality). Therefore, a greater perceived relative deviation of the OG from the prototype of the SOG compared to the IG justifies therefore negative attitudes toward the OG (Waldzus, Mummendey, Wenzel, & Weber, 2003). An application of the Ingroup Projection Model to one of the initial examples may illustrate its implications. Workers in general constitute a common SOG for both striking (IG) and non-striking workers (OG). Striking workers project their IG-goal of striking for better wages to the SOG of all workers. The OG-behavior of not striking is evaluated with reference to this representation of the SOG goal of striking. The deviance of the non-striking workers (OG) is evaluated in relation to the IG-goal of striking that is projected on the SOG. The OG's perceived deviance from this SOG-goal justifies a negative evaluation of the OG and even negative behavior towards its members.

The current research is concerned with determinants of the extent of harm that is inflicted on deviant groups. Why are workers who decide not to join a strike effort sometimes merely frowned upon by their fellow workers and at other times threatened with physical violence and even murder? The following chapter addresses how the very same act of deviation may lead to different levels of harm inflicted on a group of deviants.

2.2 Minimal and Maximal Goals

According to Carlsmith, Darley, & Robinson (2002), people tend to punish relative to the degree of deviance and to the perceived maliciousness of intention of a deviating individual or group. They showed in several studies that the factors degree of deviance and non-normative intent determine the degree of punishment. Other factors related to utilitarian ideologies - such as a deterrence motive - have only a smaller effect on the extent of assigned punishment (see also Darley, Carlsmith, & Robinson, 2000). But why do people perceive some deviations of an OG as only relatively negative while other people perceive the same deviations as absolutely negative? Mummendey and co-authors proposed two basic types of goals that differ fundamentally with respect to the evaluation of an event relative to that goal: Minimal and maximal goals (Mummendey, Waldzus and Kessler, 2002; see also: Kessler, Neumann, Mummendey, Berthold & Waldzus, 2010). Maximal goals are desired or undesired end states that can be approached or avoided. Events which are evaluated relative to maximal goals are judged by their relative distance from these goals and therefore receive a graded evaluation. The valence of an event relative to a maximal goal depends on its distance to the goal. In contrast to this "shades of grey" type of evaluation, minimal goals lead to "black/ white" evaluations of an event. Minimal goals define a certain point that must be exceeded. A minimal goal can therefore only be fully met or not met at all. An event which is evaluated relative to a minimal goal receives a dichotomous evaluation and is therefore either entirely positive or negative. It is assumed that the representation of a superordinate goal[1] as either minimal or maximal not only affects the evaluation of an event relative to this goal, but also has strong implications on the behavioral reactions towards members of an OG that is deviating from this goal. The deviation from a minimal goal should lead to more negative reactions compared to the deviation from a graded maximal goal, because any deviation is evaluated as completely negative regardless of the absolute degree of deviation. Several

[1] The term „superordinate goal" will be used in the following to refer to the group goal of a superordinate group.

studies by Kessler, Neumann, Mummendey, Berthold, Schubert, and Waldzus (2010) and Fritsche, Kessler, Mummendey, and Neumann (2009) support the distinction between minimal and maximal goals and provide evidence for their specific effects on the evaluation of deviants. Kessler et al. (2010) showed that the degree of deviance and the type of standard - minimal versus maximal - interact with regard to the level of assigned punishment behavior. The degree of deviance did not affect the severity of the assigned punishment for participants oriented towards a minimal standard. In contrast for participants oriented towards a maximal standard the degree of deviance was crucial for the level of assigned punishment. Fritsche et al. (2009) assessed quasi-experimentally how individual chronic differences in the tendency to set minimal versus maximal goals increase punishment inclinations and social exclusion tendencies. Berthold, Mummendey, Kessler, and Lücke (submitted) showed further that the fundamental differentiation between minimal and maximal goals applies as well to intergroup relations. They could show in the framework of the Ingroup Projection Model (Mummendey & Wenzel, 1999) that minimal goals are of specific importance for ingroup projection and therefore affect attitudes towards OGs deviating from superordinate goals. It is the aim of the present research to extend these findings in an experimental setting by showing the impact of minimal and maximal representations of a superordinate group goal on the infliction of harm on members of an OG deviating from this superordinate goal. The following chapter sets out to integrate this research objective in the broader framework of existing research on social discrimination.

2.3 The Infliction of Harm on Deviants

2.3.1 Social Discrimination

An extensive body of research dealing with social discrimination has done so focussing on stereotypes of and prejudices towards OGs. However, the present research is explicitly concerned with the negative *behavioral* manifestations of social discrimination. It is focusing on the phenomenon of blatant negative discrimination within the broader category of social discrimination: The intentional infliction of harm on members of a deviant OG (e.g. Bar-Tal, 1989; Bar-Tal, 1990, Bar-Tal, 2004, Staub, 1989, 1999). The central research question is concerned with the extent of harm that is inflicted on group members deviating from a common goal. What is it that determines if these deviants are punished weakly or severely? Why are strike-breakers sometimes merely frowned upon by their fellow workers and at other times murdered?

More than 50 years ago Sherif developed his theory of realistic group conflict (Sherif, White, & Harvey, 1955; Sherif, Harvey, White, Hood, Sherif, 1961). The theory was based on the results of the summer camp experiments dealing with group formation, the development of intergroup conflict and the possibility to resolve this conflict. The Realistic Group Conflict Theory states that attitudes as well as behavior toward members of IG and OG are consequences of the functional relation between the goals of IG and OG. An intergroup relation characterized by conflicting group goals - negative interdependence - leads according to the Realistic Group Conflict Theory to a perception of competition with the OG, prejudice and derogative behavior toward the OG and at the same time to a positive evaluation of one´s IG. Subsequent research could show that a negative interdependence of IG and OG goals was not an essential precondition to elicit discriminatory tendencies. A common fate of members of a group is an important factor in the equation leading to IG favoritism and OG derogation (Rabbie & Horwitz, 1969; Horwitz & Rabbie, 1982). This function was further specified by Tajfel and co-authors and the development of the Minimal Group Paradigm. The initial motivation for the development of the Minimal Group Paradigm was the definition of a baseline condition for no discrimination to occur (Tajfel, Billig, Bundy, & Flament, 1971). First results showed that the mere categorization in salient but trivial groups was sufficient to elicit IG favouritism. It was specifically observed that group members were using a strategy to maximize the relative differences between their IG and an OG even at the cost of sacrificing the maximum profit for their IG (*maximizing difference strategy*). A rich body of subsequent

research confirmed this initial finding on a multitude of different measures (e.g. Brown, 2000; Messick & Mackie, 1989). Social discrimination in the form of IG favouritism can occur as a result of a mere group categorization. Social Identity Theory and Self-Categorization Theory developed as theoretical foundations for these results that emphasise a comparative perspective (Tajfel & Turner, 1979; 1986; Turner, Hogg, Oakes, Reicher & Wetherell, 1987). This approach builds on peoples´ motivation to sustain a positive identity. In a situation in which people define themselves on the basis of a salient social identity, a positive evaluation can be established by a comparison outcome between their IG and a relevant OG that creates positive distinctiveness for ones IG in relation to the OG. In both Social Identity Theory and Self-Categorization Theory discrimination against an OG is seen as an individual´s attempt to achieve positive distinctiveness via its IG identity in a salient intergroup context (Mummendey & Otten, 1998). Maximizing the distance between IG and OG on a relevant dimension of comparison should - according to this comparative perspective on discrimination - not be restricted to a positive domain but also hold true in a negative domain (see Table 1). Empirical evidence however showed that IG favoritism and discrimination of the OG in a positively valued domain are not necessarily related to explicitly negative intentions to inflict harm on the OG. Struch and Schwarz presented a study focussing on intergroup aggression toward an OG of ultraorthodox Jews (Struch & Schwarz, 1989). The measures of intergroup aggression included the opposition to institutions serving the needs of the OG, the support of harmful acts toward the OG and the opposition to interactions with the OG. The perceived conflict of interests between IG and OG was highly related to aggressive intentions toward the OG. Participants showed also OG-favoritism in trait evaluations, but it was unrelated to any aggressive intentions toward the OG.

A multitude of experimental studies document systematic differences between the assignment of positive and negative resources to one´s IG and an OG. It could be shown that participants who had to assign an aversive resource - varying durations of an unpleasant sound - to IG and OG members did not show IG favoritism or follow a maximizing difference strategy. Instead fairness was the dominant strategy for the assignment of this negative resource to members of the arbitrarily assigned IG and OG (Mummendey, Simon, Dietze, Grünert, Haeger, Kessler, Lettgen, & Schäferhoff, 1992). Mere identification as established in the Minimal Group Paradigm alone appears to be insufficient for explicit negative OG derogation to occur (Mummendey & Otten 1998). The general effect that IG favoritism is less likely to occur in the domain of negative resource allocation than it is in the domain of positive resource allocation has been coined the *Positive Negative Asymmetry of Social*

Discrimination. The central finding of the *Positive Negative Asymmetry* was supported by a meta-analysis sampling 52 studies investigating in-group bias using an experimental manipulation of valence of the experimental condition (Buhl, 1999). Participants discriminated in these studies significantly less on negative scales than on positive scales. The effect sizes from the positive and negative valence conditions were not significantly correlated suggesting that different processes underlie intergroup discrimination in a positive respectively a negative domain.

Table 1: A Taxonomy of Social Discrimination (from: Mummendey & Simon, 1991; Mummendey & Otten, 1998)

	Valence			
Behavior	*Positive stimuli*		*Negative stimuli*	
Allocation	Direct discrimination	IG > OG (a)	Direct discrimination	IG < OG (b)
Removal	Indirect discrimination	IG < OG (c)	Indirect discrimination	IG > OG (d)

Attempts to explain the Positive Negative Asymmetry included valence specific differences in information processing (Mummendey & Otten, 1998) and the effect of stimulus-valence on the inclusiveness of self-categorization (Mummendey, Otten, Berger & Kessler, 2000). It could be shown in a motivational approach to the Positive Negative Asymmetry that the allocation of positive and negative resources depends on the match of regulatory focus and the valence of the distributed resources (Sassenberg, Kessler, & Mummendey, 2003). Social discrimination was higher when social identity was salient and a regulatory fit between group based regulatory focus and valence of distributed resources was given. It could be shown that a simple framing lead to a reversal of the *Positive Negative Asymmetry* for participants with a momentary prevention focus. Social discrimination occurred under a promotion focus when a positive resource was allocated and under a prevention focus when a negative resource was allocated. No such effect was found for the distribution of negative resources under a promotion focus or for the distribution of positive resources under a prevention focus.

The present research is in contrast to the research on the *Positive-Negative Asymmetry of Social Discrimination* not concerned with minimal conditions for discrimination based on negative resources to occur. This line of research is instead focusing on the degrees of harm inflicted on members of a deviant OG (Struch & Schwarz, 1989, Levin & Sidanius, 1999). But what are then "aggravating" conditions leading to explicit negative discrimination of members of an OG (s. Table 1)? Brewer (Brewer, 1999; 2001) criticized the idea of a direct mutual conditionality of IG favoritism and explicit negative OG derogation as initially stated by Sumner: "Loyalty of the group, sacrifice for it, hatred and contempt for outsiders, brotherhood within, warlikeness without - all grow together, common products of the same situation" (Sumner, 1906). Instead Brewer suggested a conceptual separation of favoritism of one´s own group over another, "Ingroup Love", on the one side and "Outgroup Hate", the explicit derogation of an OG, on the other (Brewer, 1979; Brewer, 2001). Brewer argues that OG hostility is based on social categorization and identification with a group but that these factors are not sufficient and that this relation is by no means causal. She proposes rather that OGs might become relevant in relation to IG based motives such as the need for secure inclusion, the protection of IG distinctiveness (Brewer, 1991) and a positive evaluation of one´s IG in relation to other groups. The competition for social status and the perception of an OG as a potential threat with regard to one of these motives serves as a breeding ground for hostilities toward an OG. This line of argument coincides with research conducted on the *Positive Negative Asymmetry of Social Discrimination*. Mummendey and co-authors proposed an *Aggravation Hypothesis* on the basis of the *Positive Negative Asymmetry of Social Discrimination*. Social discrimination in a negative domain requires according to this hypothesis not only a salient social categorization but additional factors that exacerbate the pursuit of a positive group-based distinctiveness. Research by Blanz, Mummendey, and Otten showed that social discrimination in the negative domain occurred only when two aggravating conditions co-occurred, namely a high salience of size- and status- similarity between IG and OG and a high level of IG identification (Blanz, Mummendey, & Otten, 1995a, 1995b). It was argued that only individuals who are motivated to strive for a positive social identity - because of their threatened or negative social identity - assign more negative resources to an OG than to their own IG (Mummendey & Otten, 1998).

Another line of research has identified various forms of intergroup threat as important factors leading to negative OG attitudes (Riek, Mania, & Gaertner. 2006; Stephan & Stephan 1996). Sherif´s aforementioned Realistic Group Conflict Theory constituted an important starting point for this research. Conflicting group goals lead according to the Realistic Group

Conflict Theory to higher levels of perceived intergroup competition and more negative attitudes toward the OG (e.g. Esses, Dovidio, Jackson, & Armstrong, 2001; Esses, Jackson, & Armstrong, 1998). This is even the case if individual group members are not directly affected by the outcome of the intergroup conflict (Bobo, 1983). This approach has been complemented by alternative approaches to intergroup threat. Symbolic threats such as conflicting values of groups may lead even in the absence of any realistic intergroup conflict to an increase in negative attitudes toward members of an OG (Esses, Haddock, & Zanna, 1993; Kinder & Sears, 1981, Sears 1988). More recently Stephan and Stephan proposed their *Intergrated Threat Theory* in an attempt to integrate different theoretical approaches on intergroup threat including realistic threat (1), symbolic threat (2), and negative stereotypes (3) but also threats to group esteem (4)and intergroup anxiety (5) (Stephan & Stephan, 2000). An increase in perceived intergroup competition with an OG, value violations and group esteem threat by the OG promote, according to the *Integrated Threat Theory*, intergroup anxiety and the importance of stereotypes which in turn entail an increase in negative attitudes toward the OG (Riek, Mania, & Gaertner, 2006; Stephan & Stephan 1996).

This chapter reviewed a selection of the social psychological literature that is highly relevant to the phenomenon of social discrimination. It was the main goal of this review to clearly define the specific aspect of social discrimination that the present research is focusing on: The intentional infliction of harm on members of a deviant OG. The following chapter introduces a second, quite different theoretical perspective on the infliction of harm on deviants. This research has so far been essentially unconnected with the presented social psychological literature. It focuses on research in the field of experimental economics that is largely concerned with the importance of cooperation in groups. The question is raised whether the infliction of harm on members of deviant subgroups might be interpreted as a process maintaining a group´s functionality and continuity.

2.3.2 The Infliction of Harm on Deviants: Experimental Research on Punishment in Public Goods Games

A diverse body of research in experimental economics has been dealing with the infliction of harm on deviants and its functional implications. In this line of research in experimental economics the question why humans punish deviants is closely related to the phenomenon of cooperation. Human cooperation is distinct from cooperation in other animal species which is mostly restricted to cooperation among genetically related individuals (Hamilton, 1964; Fehr

& Fischbacher, 2004a). Humans however developed a qualitatively different kind of cooperation: Cooperation among genetically unrelated individuals (Henrich & McElreath, 2003). It might be speculated that the emergence of cooperation among genetically unrelated group members was intertwined with the increasing importance of ever larger groups. Group sizes exceeding families and village communities required new forms of cooperation. At the same time these new forms of cooperation assured the functionality of large groups of genetically unrelated individuals such as nations that would not have been feasible on the basis of mere kin-selection. But cooperation itself is a fragile phenomenon. Under which circumstances do members of a genetically unrelated group cooperate toward a common group endeavor? It might be argued that many group endeavors that require the cooperation of its group members might be conceptualized as Public Goods Games (e.g. Kollock, 1998; Pruitt & Kimmel, 1977).

2.3.2.1 Public Goods Game

The initial example of protecting one´s country in a war constitutes a classic example of such a common group endeavor. The citizens of a country are faced with the decision to fight for their country or to defect. Fighting for one´s country is potentially costly as it implies the possibility of being wounded or killed in battle. Therefore, not participating in a war should be the preferable decision from an individual standpoint. However, if all citizens decided not to participate in the effort to protect their country the war would be lost. This should result in the second-worst outcome for citizens from an individual perspective - only second to being killed - as there is no payoff from the common group endeavor. Their country did not win the war but was instead conquered. On the upside they did not participate in the war and were not faced with the danger of being wounded or killed in battle.

The reasoning behind this example of a common group endeavor can be conceptualized as a more formal Public Goods Game. In a Public Goods Game a group of participants is formed ($n \geq 2$). All players[2] are given an initial endowment e. Participants are then simultaneously faced with the decision of how much of this initial endowment they want to keep for themselves and how much they want to contribute to the Public Good. The sum of all contributions is then multiplied by a factor m that is larger than 1 and smaller than n. The product of all contributions and m is the total payoff to all participants. The total payoff is distributed equally among all players, regardless of their contribution. Every participant

[2] Individuals participating in a Public Goods Game will furthermore be referred to as “players”. Individuals participating in a study will be referred to as “participants”.

receives m/n monetary units for every contributed monetary unit, $m/n < 1$ because $m < n$. This incentivizes free-riding. The individually rational decision would be to defect, to contribute nothing. If the other players contributed, a free-rider would receive a combined payoff of the kept initial endowment and his share of the total payoff from the Public Good (1). If all other players defected as well the free rider would receive at least the entire initial endowment (2). In both scenarios - (1) and (2) - the free-rider could not improve his payoff by changing his contribution decision. The collectively rational decision would be to contribute the entire initial endowment. If all players contributed their initial endowment the collective payoff, then $n*(m*e)$ would be larger than the collective payoff if all players defected and kept their initial endowment $n*e$. A Public Good Game constitutes a Social Dilemma because of this antagonism of individual and collective rationale.

A large body of research emphasizes the importance of specific social norms for the emergence and upkeep of cooperation in Social Dilemmas (e.g. Fehr & Gächter, 2004). One social norm prescribing cooperation, if the other members of a group cooperate as well, is the norm of conditional cooperation. Experimental evidence shows that many individuals tend to behave in a conditionally cooperative manner (e.g. Fehr & Gächter, 2004). Many of these conditional co-operators are more precisely imperfect conditional co-operators as they contribute slightly less than other co-operators. This indicates that norm adherence is curtailed by individual self interest. Another stable subpopulation of players in Public Good Games behaves in an individually rational manner and defects consistently. Both deviations from the conditional cooperation norm lead to a decreasing level of cooperation over time (Dawes & Thaler, 1988; Ledyard, 1996). But the introduction of two factors curbs the escalation of defection and seems to facilitate a high level of cooperation: The identifiability of individual contribution decisions and the possibility to punish non-contributors. A broad body of research has been able to show the positive effects of specifically targeted punishment on the level of cooperation in Public Good Games (Ostrom, 1992, 2000; Yamagishi, 1986; 1988; Fehr & Gächter, 2000; 2002). The introduction of a punishment option results in these studies not only in an immediate increase in cooperation, but also in a further increase in cooperation over time, thus reversing the cooperation pattern over time in a Public Good without a punishment option. The impact of a punishment option is twofold from a motivational perspective: First, the expectation that deviants will be punished creates a perspective of a high level of cooperation for potential co-operators and may be best described as an "attraction of cooperation" (1). The possibility of an impending punishment poses a threat to potential defectors and decreases the incentive to free-ride. This second implication of a

punishment option can be described as a "suppression of competition" (e.g. Price, Cosmides, & Tooby, 2002; Frank, 2003) (2). An application to the initial example of fighting of one´s country may illustrate these implications of a punishment option in a Public Goods Game.

2.3.2.2 Punishment and Cooperation in Public Good Games: An Example

The protection of one´s country in a war has been introduced earlier as a typical group endeavor. However, individual defection constitutes a serious threat to this group goal. The introduction of a punishment option has been presented as a means of securing a high level of cooperation towards the group goal of protecting one´s country. Joining the army in case of a draft resembles cooperation with regard to this group goal. Desertion would amount to an act of defection with regard to this group goal. The individually rational decision would be to defect and "keep ones initial endowment". This would mean in terms of the example to minimize the probability of being killed in an armed conflict and stay at home. At the same time one would hope that all other group members contributed and one´s group would win the conflict. If however all group members would follow this individual rationale the conflict would be lost, resulting in no additional payoff from the Public Good for anybody. The collectively rational decision would be to join the army and maximize the probability of winning the conflict. Given the fact that the initial endowment is the ultimate contribution of risking one´s life, traditional punishment options such as executions have to be equally severe to counterbalance the cost of contribution. The initial endowment in this example is not a positive amount of a positively valued resource but the probability p of a negative event not taking place, $1 - p(d)$, ones death d is assigned a set value of 1, the probability of this event p ranging from 0 to 1. The probability of being killed while staying at home equals 0 in this example. The payoff for defecting is therefore equal 1 if no punishment option exists. In the case of a punishment option in form of executions of deserters the payoff for defection is set to 0 by definition because death is certain in this case, $p(d) = 1$. The risk of being killed in battle is larger than 0. But it is still smaller than the certain execution in case of defecting. The individual payoff for contributing for this given punishment option is therefore larger than for defecting, $1 - p(d) \geq 0$, for any given p. In other words: Going to war with any probability of being killed is still preferable over being killed at home for sure as punishment for ones desertion. This implication of a punishment option has been described earlier as the "suppression of competition" (2). The decline in desirability of defection in case of a punishment option increases at the same time the expected probability of a high level of cooperation and thus a high additional payoff from the Public Good. The more individuals

decide that desertion is not worth the risk of being executed, the more individuals join the army and thus increase the probability of their country to win the war. This function of a punishment option has been described earlier as the "attraction of cooperation" (1).

An option of punishing deviants may appear appealing with regard to establishing a high level of cooperation in a group. However, punishment is costly. This cost does not have to be of a monetary nature. But punishment requires the investment of resources such as time and energy. Punishment entails at the same time the risk of retaliation and emotional tensions. Norm deviations are in addition not always blatant. The detection of any norm deviation - respectively the individual or group committing it - requires time, information, money and effort. The notion that punishment is costly has another important theoretical implication. If punishment is assigned with a cost, punishment has to be limited in quantity and quality. A first question addresses the quantity of the punishment. The previously mentioned evidence illustrating the potential of punishment to increase levels of cooperation suggests that punishment is motivated by self-interest. Strict punishment of deviations improves the level of cooperation in the Public Good Game and results therefore in a higher individual payoff. Experimental evidence has lead to a perspective on motivations for punishing behavior that is deviating considerably from this self-interest perspective. Fehr and Fischbacher reported several studies on third-party punishment in line with the findings of Carlsmith and co-authors (Fehr & Fischbacher, 2004a). In a third-party punishment experiment a third person observes two others playing a Prisoner´s Dilemma Game. A Prisoner´s Dilemma Game is basically a variation of a Public Good Game with two players that have the dichotomous decision between contribution and defection. As in the Public Goods Game defection is the individually rational and contribution the collectively rational choice. In a third-party punishment experiment an uninvolved person is first informed about the contribution decisions of the two players and then confronted with the option of costly punishment. There is not only no possibility for the third-party observer to profit by punishing one of the two players. Even more, the observer´s self-interest should obstruct the punishment of one of the two players because punishment is costly. This is however not the case. Fehr and co-authors could show that a considerable share of about 60 % of the observers punished violators of a distribution norm (Fehr & Fischbacher, 2004b).

This theoretical and empirical research suggests in summary that the punishment of deviants might be seen as functional from the perspective of a superordinate group[3].

[3] A goal of a superordinate group will hence be referred to as "superordinate goal".

Punishment incentivizes members of a group not to defect but to contribute to a group endeavor. The punishment of deviants is not only motivated by an individual´s self-interest utilizing it´s group as a mere proxy for the pursuit of its individual goals. The punishment of deviants might be seen as a measure to ensure a group´s functionality and thus preserve its continuity.

2.3.2.3 Social Norms and Group Goals

Fehr and Fischbacher define social norms as "[...] standards of behaviour that are based on widely shared beliefs how individual group members ought to behave in a given situation" (Fehr & Fischbacher 2004a). This particular definition of social norm is somewhat misleading as it is targeted at the particular "norm of strong reciprocity". However, strong reciprocity might be better described as a strategy of norm enforcement. Goals on the contrary can be broadly defined as internal representations of desired states (Austin & Vancouver, 1996). A group goal is accordingly defined as an internal representation of a desired end state that is a part of a specific social identity. Social norms are sets of values, beliefs, attitudes and forms of behavior that are perceived as appropriate or inappropriate with regard to a group goal. The existence of a valid social norm implies that the group goal is shared by other members of the group. A socially shared social norm suggests even more importantly the possibility of negative social sanctions and therefore increases the probability of other group members´ norm compliance. At this stage strategies of norm enforcement - such as strong reciprocity - are highly crucial as they connect individual behavior with a relevant social norm that is implicitly or explicitly derived from a group goal. Strategies of norm enhancement incentivize individuals to behave in line with the "do´s and don'ts" that are defined by a specific social norm. The definition of desirable form of values, beliefs, attitudes and forms of behavior by a social norm again is most important for the achievement of a group goal. This perspective is in line with a functional approach to normative behavior that argues that social norms aid the accomplishment of group goals (Allison 1992; Campbell, 1975, Sherif, 1936; Sumner, 1906). Norms are according to this perspective neither arbitrary nor trivial but ultimately connected to the survival of a group in their capacity to encourage or prevent specific types of behavior. Social norms that do not aid any ultimate goal are not expected to be evolutionary stable (Schaller & Latane, 1996). Both group goals and social norms are important components in common group endeavors. Group goals define to a greater extent the "what", the direction of the group goal and its motivational implications whereas social norms define the "how".

Strategies of norm enforcement such as strong reciprocity are pivotal as they ensure that individuals behave in line with specific norms, thus contributing appropriately to the common endeavor of the group goal.

An example might illustrate this differentiation: Workers are defined as a group by the common group endeavor to negotiate better working conditions and higher wages. In order to achieve this group goal the group members participate in strikes or protest marches. This group goal might be described as a Public Goods Game in which a goal achievement, represented by higher wages and better working conditions, is symbolized by the additional payoff from the Public Good. Group members have to contribute shares of their initial endowment in order to achieve this group goal. In real life striking workers may contribute to the Public Good of the group goal by costly investments in terms of time spent on strike, money missed out on because of the missed working opportunities. The social norm would be to contribute to the group goal of striking by individuals investing these resources. Both striking workers as well as non-striking workers are included in the superordinate social category of workers even though they show different contribution patterns. The defection by a deviant group - such as strike-breakers - would be punished as it undermines the common group goal.

This example might not only help to clarify some differences in terms of the used terminologies, but furthermore show that the two different theoretical approaches might be connected. From an OG-centric perspective the strike-breakers are discriminated as they are punished by the strikers for their perceived defection. The perceived interdependence (Esses et al. 1998; Esses et al., 2001) and possible negative implications for the success of the strike might serve as "aggravating" conditions justifying the infliction of harm on the members of this deviant OG[4]. However, the situation might be interpreted as an IG internal conflict from a functional perspective. The strikers have to punish defectors in order to keep a high level of cooperation and thus maximize the chances of group goal achievement.

2.3.3 The Infliction of Harm on Deviant Groups: Summary

Two different bodies of literature have been presented as a theoretical foundation of the present research. Both lines of research are highly relevant with regard to the infliction of harm on deviants. While *social psychology* has traditionally taken a rather descriptive

[4] It is important to point out that no structural negative interdependence exists between both groups as a Public Goods Game constitutes a mixed-motive situation.

approach focussing on processes that might help explain the negative epitomizations of this phenomenon such as explicit discrimination and extreme forms of persecution and punishment of deviance in cases such as ethnic cleansings. *Experimental economics* - and evolutionary psychology - on the contrary have focussed largely on a normative perspective on the punishment of deviants. The punishment of deviants is often viewed as a mediating factor suited to increase levels of cooperation in groups. These two approaches vary at the same time with regard to the perspective they take on the infliction of harm on deviants and its subsequent evaluation. *Social psychology´*s view is often characterized by the perspective of a deviant OG that is confronted with being discriminated. The normative approach of *experimental economics* on the contrary focuses on these functional aspects of the infliction of harm on deviant. Punishment is seen as a factor that is crucial for the upkeep of cooperation in a group. It is nevertheless most important to emphasize that both lines of research are different approaches dealing with processes related to the same phenomenon, the infliction of harm on deviants. Both approaches may differ in terms of their methodology, terminology and general perspective, but they are far from being incompatible.

The current research is connected to both approaches to the infliction of harm on deviants as it focuses on the *extent of harm* that is inflicted on members of a deviant group. It has been suggested that the representation of a superordinate goal as either minimal or maximal goal is crucial for the extent of harm that is inflicted on members of a group that is deviating from this goal. An empirically testable hypothesis is therefore derived from this theoretical assumption.

2.4 Hypotheses

2.4.1.1 Hypothesis 1: Goal-type Hypothesis

It is expected that the deviation from a minimal superordinate goal leads to a higher level of harm inflicted on members of a deviant OG than the deviation from a maximal superordinate goal.

2.4.1.2 Negative Emotions

The infliction of harm on others is often triggered by emotional reactions such as moral outrage (e.g., Carlsmith, Darley, & Robinson, 2002). An affective level of analysis appears therefore important in order to scrutinize the impact of goal-type on the extent of harm that is inflicted on members of a deviant OG. An important set of negative emotions has been

presented with the CAD triad hypothesis (Rozin, Lowery, Imada & Haidt, 1999). Rozin and colleagues propose a set of three negative emotions – contempt, anger and disgust (CAD) – and suggest that each of them is related to a specific moral code. Although distinguishable from each other these three emotions can generally be classified as a cluster of emotional reactions to moral violations. The deviation from a minimal goal should lead not only to a higher level of harm inflicted on members of a deviant group than the deviation from a maximal goal (*goal-type hypothesis*). The dichotomous structure of a minimal superordinate goal should also lead to a higher level of negative emotions toward the deviant OG than the deviation from a graded maximal goal. The higher level of negative emotions toward the OG should in addition be related to the behavioral reaction - the infliction of harm - thus mediating the main effect of goal-type of the superordinate goal on the infliction of harm on members of the deviant OG.

2.4.1.3 Hypothesis 2: Negative Emotions Hypothesis

It is expected that negative moral emotions toward the OG mediate the effect of goal-type of the superordinate goal on the degree of harm inflicted on members of the deviant OG.

In the following chapter an experimental paradigm will be developed in order to test both *goal-type hypothesis* as well as *negative emotions hypothesis*.

3 Designing an Experimental Paradigm

The basic question of the present research addresses the impact of the goal-type of a group goal (minimal versus maximal) on the infliction of harm on a subgroup that is deviating from this goal. An appropriate experimental paradigm is needed to answer this question. This experimental paradigm should incorporate a number of important methodological aspects. First of all it should allow the introduction of a *superordinate goal*[5]. Secondly, *an intergroup context* has to be introduced to allow the *deviation of a subgroup from this superordinate goal.* The experimental paradigm should allow the behavioral measurement of *harm inflicted on the deviant subgroup* in response to the deviation of this subgroup from the superordinate goal. The experimental paradigm has to feature a *goal-type* (minimal versus maximal) *manipulation of the superordinate goal* in order to assess the hypothesis that the deviation from a minimal superordinate goal leads to a higher level of harm inflicted on deviants compared to the deviation from a maximal superordinate goal. The entire experimental paradigm should be based on an *experimental setup* in order to exclude the influence of uncontrolled factors. The following chapters will address each of these desiderata and discuss how their practical operationalizations may be incorporated into a coherent experimental paradigm.

3.1 Superordinate Goal

Sherif´s aforementioned summer camp studies present a classic example of research on group goals and intergroup relations (Sherif, White, & Harvey, 1955; Sherif, Harvey, White, Hood, Sherif, 1966). The Theory of Realistic Intergroup Conflict that was derived from these findings states that a negative interdependence between the goals of an IG and an OG results in negative attitudes and behavior towards the OG. The present research is in contrast not concerned with intergroup conflict on the basis of a negative interdependence between two groups. The focus of the present research is instead the infliction of harm on a subgroup that is deviating from a group goal that is potentially benefitting all group members. A Public Goods Game allows the introduction of such a group goal without the necessity of a zero-sum conflict. A Public Goods Game represents a basic form of a mixed-motive game. The

[5] The term „superordinate goal“ will be used in the following to refer to the group goal of this superordinate group.

essential characteristics of a Public Goods Games have already been described earlier (see chapter 3). Players in a Public Good Game receive an initial endowment and are faced with the decision of how much of that initial endowment they want to contribute to a Public Good. The collectively rational choice would be to contribute the entire initial endowment to the Public Good as the Public Good benefits all players. It has been argued (see chapter 3) that every group goal may be conceptualized as a Public Goods Game. The Public Good - respectively the additional payoff from it - constitutes the group goal[6]. The present research is dealing with the infliction of harm on members of a deviant subgroup. The next section deals therefore with the introduction of an intergroup context: The notion that the superordinate group consists of members of one´s own group but also of members of a distinct OG.

3.2 Intergroup Context: Social Identity

The point of departure for the present research is a group striving for a common group goal. A subgroup is however deviating from this group goal, thus introducing the distinction between an ingroup (IG) that is conform to this superordinate goal[7] and an outgroup (OG) that is deviating from this superordinate goal. Both IG and OG are included in a superordinate group (SOG) (Figure 1). It appears however reasonable from a methodological point to introduce this distinction between an IG and an OG included in a SOG prior to the actual deviation of the OG from the superordinate goal. These group categories may function on the event of the OG´s deviation as a "predetermined breaking point". Introducing the intergroup context prior to the OG´s deviation should give participants greater clarity about the experimental situation and function at the same time as a basis for the attribution of the OG´s deviation ("Those who are different by definition are the ones that deviate from our common group goal."). It is highly important to point out that this intergroup context should have no functional implications for the group goal. The intergroup context should also be as content-

[6] *Contribution Decisions as an Indirect Manipulation Check of Group Goal*. The contribution decision functions at the same time as a logical manipulation check for the goal-manipulation. A Public Goods Game is a case of a Social Dilemma. It does not make sense to contribute if a player does not believe that the other players will do the same. It only makes sense to contribute ones initial endowment in order to maximize ones payoff if one is convinced that the other players will do the same. A player´s decision to contribute to the Public Good indicates therefore not only the representation of a shared group goal of making money by contributing to the Public Good. It also indicates the expectation that other players will behave in line with the group goal of contributing to the Public Good.

free as possible to avoid the influence of factors such as stereotypes or the interaction history between the groups.

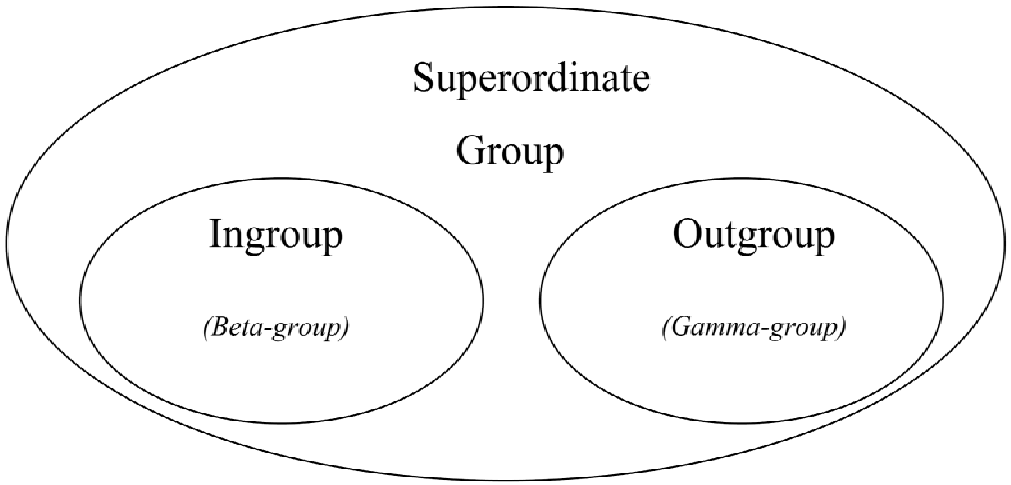

Figure 1: Group Categorizations

The Minimal Group Paradigm constitutes a prominent experimental baseline situation for a context-free intergroup situation (Tajfel, Billig, Bundy, & Flament, 1971). Tajfel and co-authors wanted to demonstrate that social categorization in itself is no sufficient condition to elicit intergroup discrimination in the form of IG favoritism. Six basic criteria were proposed for their Minimal Group Paradigm (Tajfel et al., 1971; Diehl, 1990):

1. There should be no face-to-face interaction whatever between the Ss, either in the ingroup or in the outgroup or between the groups.
2. Complete anonymity of group membership should be preserved.
3. There should be no instrumental or rational link between the criteria for intergroup categorization and the nature of ingroup and outgroup responses requested from the subjects.
4. The responses should not represent any utilitarian value to the subject making them.
5. A strategy of responding in terms of intergroup differentiation (i.e. favouring the ingroup and detrimental to the outgroup) should be in competition with a strategy based on other more “rational” and “utilitarian” principles, such as obtaining maximum benefit for all. A further step in this direction would be to oppose a strategy

of maximum material benefit to the ingroup to one in which the group gains less than it could, but more than the outgroup.

6. Last but not least, the responses should be made as important as possible to the Ss. They should consist of real decisions about the distribution of concrete rewards (and/or penalties) to others rather than some form of evaluation of others.

The following paragraph describes how these criteria of the Minimal Group Paradigm were adapted to the experimental paradigm of the present research. The operationalization of the Minimal Group Paradigm is identical for all studies. In the beginning of the main study participants are welcomed and taken to individual sound-proof booths. After the study participants are thanked and paid individually. There is no face-to-face contact or any kind of communication between the participants of the studies at any time during the study (*criteria 1 and 2*). An intergroup context is introduced using a brief pretest during the recruitment of the participants several days in advance of the actual main study. The questionnaire consists of ten questions taken from the Freiburg Personality Inventory (Fahrenberg, Hampel, and Selg, 2001) in a random manner (see Appendix: Group Categorization). These questions are neither related to each other nor to the subjective of the following main study. The group categories are therefore arbitrary. Participants are allegedly classified as members of one of two groups, a "beta-group" and a "gamma-group". These groups are characterized as broad personality-types, but no further information is given. All participants receive feedback that they have been classified as members of the "beta-group" (*criterion 3*). Both IG ("beta-group") and OG ("gamma-group") are included in a superordinate group (Figure 1). Participants cannot profit from any of their decisions on the dependent variables of the study (*criterion 4*). The decisions that are based on the group categories are comprised either of strong intentions (e.g. "A certain amount of money should be deducted from the OG´s payoff") or actual behavioral choices (e.g. "0 - 9 € should be deducted from the OG´s payoff"; "Members of the OG should be excluded from the participation in future profitable studies.") (*criterion 6*). These decisions are however not in competition with a strategy based on "rational" or "utilitarian" principles (*criterion 5*). The present research is - in contrast to the initial research in the framework of the Minimal Group Paradigm - focusing on the explicit infliction of harm on members of a deviant OG. The aims of these two lines of research are therefore clearly distinct from each other. The use of the Minimal Group Paradigm in the current research is purely methodological as it constitutes a well established context free technique to introduce an intergroup context.

It is important to point out that the introduction of minimal group categories has no functional implications in terms of interdependence between IG and OG. The functional relation between all participants is structured by the Public Good Game. All players are equal in this Public Good Game regardless of their group membership. The intergroup context introduced on the basis of the Minimal Group Paradigm serves simply as a "predetermined breaking point" on the onset of the OG´s deviation from the superordinate goal. The next section describes therefore the operationalization of the OG´s deviation from the described superordinate goal.

3.3 Goal Deviation

The operationalization of an OG deviation has to fulfil two basic criteria. The deviation from the superordinate goal has to be perceived to be directly related to group membership: The goal deviation should be identical across both minimal and maximal goal condition. All IG members should behave in line with the superordinate goal and all OG members should deviate from it (Figure 2). This poses a problem as group membership is based on the criteria of the Minimal Group Paradigm. It implies that there is "no instrumental or rational link between the criteria for intergroup categorization and the nature of ingroup and outgroup responses requested from the subjects". There is no reason to assume that participants´ contribution decisions should vary systematically in line with their group membership. Participants´ contribution decisions should be subject to chance. The probability that the three members of the IG contribute in line with the superordinate goal and the three OG members deviate from it (see Figure 2) has to be considered very low. It is therefore the logical approach to the first criterion to stage the decisions of the other five players. The feedback about the contribution decisions of these five co-players can be manipulated in a way that the two fellow IG-players contribute at a level that is conform with the superordinate goal whereas the three OG-players deviate significantly from the superordinate goal (Figure 2). The OG-deviation could be kept exactly identical across both minimal and maximal goal condition by staging the contribution decisions of the co-players. But at the same time the question has to be raised if the use of deception is ethically justifiable in an experimental paradigm.

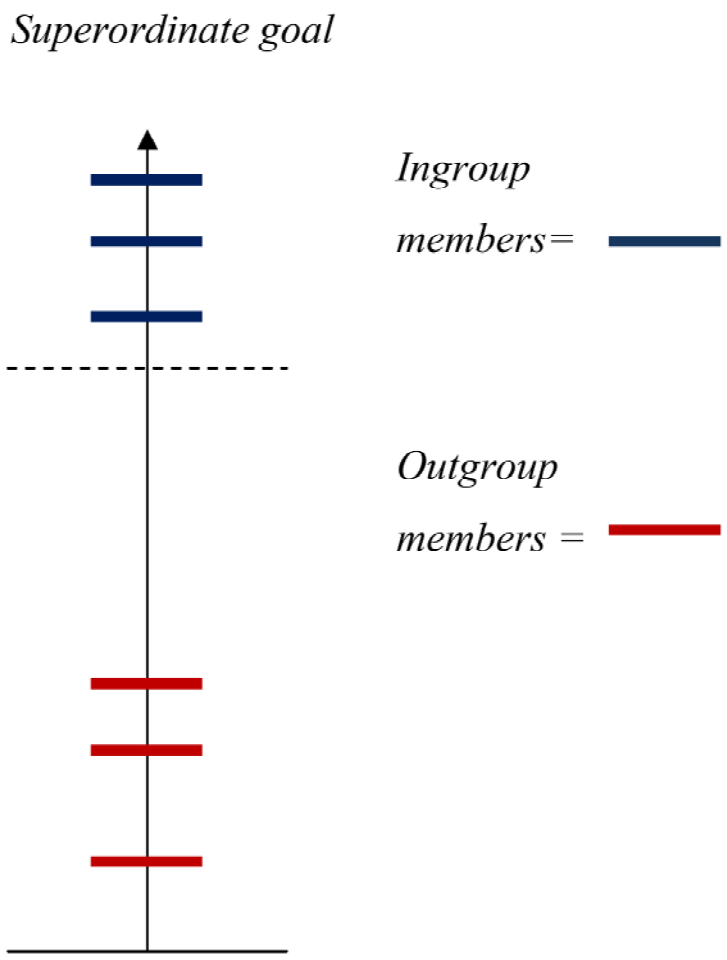

Figure 2: Goal Deviation Outgroup - Contribution Levels

3.3.1 The Use of Deception in Experimental Studies

Experimental games have been a matter of interest and research in both experimental economics and social psychology. Both disciplines do not only differ with regard to their particular research questions but also with regard to their specific ways of conducting research. Hertwig and Ortmann discuss a variety of systematic methodological differences in experimental economics and social psychology such as scripting (1), repeated trials (2) and financial incentives (3). A fourth difference has proven to be of particular sensitivity, dividing the social-scientific community into two opposing parties: The use of deception in experimental studies (4). The borders of those parties are to a large degree identical with the groups of experimental economists, who oppose the use of deception, even see it as an utter

taboo (Davis & Holt, 1993; Wilson, 2007), and the group of social psychologists who often support the use of deception - or at least do not condemn it - if it appears justified.

3.3.1.1 Contra Deception

The heated discussion can be divided along two main lines of argument, a functional and a moral one. Opponents of the use of deception in experimental studies (Hertwig and Ortmann, 2001: Davis & Holt, 1993; Hey, 1991; Ledyard 1995) often argue that participants´ responses in experimental settings should only be motivated by intended financial incentives and not by unintended reactions to presumed deception that no systematic hypotheses exist for. This point is strengthened by the concern that the negative effects of the use of deception in one study might cause spill-over effect in future studies and could affect other studies using the same population of participants. This perspective perceives participants´ trust in the investigator of an experimental study - respectively the reputation of the investigator and social scientists in general - as a common good resulting in a higher reliability of measures in experimental studies. The individual researchers "moral hazard" is from this perspective the temptation to harvest this common good without investing the effort to abstain of deception, thus endangering the continuity of the commons (Hertwig & Ortmann, 2001). The moral dimension of the argument against deception is rather plain as deception is simply perceived to be immoral and unethical.

3.3.1.2 Pro Deception

Supporters of the use of deception have a somewhat different perspective. Kimmel states for example a widespread agreement in the field of Social Psychology that deception is often a methodological necessity (Kimmel, 1996). There are two main reasons to defend the practice of defection (Hertwig & Ortmann, 2001). The investigator might lose experimental control if participants are aware of the purpose of a study. This might be especially the case in research that is targeting normatively sensible phenomena such as prejudices. Participants might adapt their responses to the perceived social norm of not showing any prejudice for example towards minorities. The authors argue that there is a greater demand in the field of (social) psychology to use deception because psychologists are more interested in phenomena that might be affected by existing social norms (Hertwig & Ortmann, 2001; Cook & Yamagishi, 2008). Secondly, deception can be used to produce situations of special interest that are unlikely to arise naturally, for example a situation that allows studying the bystander effect. The American Psychological Association proposes deception in its APA guidelines as

a last-resort strategy (American Psychological Association, APA, 2001). Many social psychologists and experimental sociologists agree that deception should only be used if necessary and should be avoided if possible (Cook & Yamagishi, 2008; Adair et al., 1985; Kelman, 1967). Surveys show that the use of deception in psychology has rather increased over the last decades (Adair, Dushenko, & Lindsay, 1985; Gross & Fleming, 1982).

3.3.1.3 Discussion

Supporters of the justified use of deception reject the moral stance that its use is immoral per se. Cook and Yamagishi turn this allegation around and suggest that it would be immoral not to conduct research that depends to a certain degree on deception (e.g. research on socially sensible issues such as xenophobia) (Cook & Yamagishi, 2008). Some proponents of the use of deception argue that some important research questions depend entirely on the use of deception (Aronson & Carlsmith, 1968). Another rather general point of criticism to the argumentation against deception in general is that there are possibly other factors besides actual deception that have the exact same consequences as deception. There is - from a psychological perspective - no reason to assume that unconfirmed suspicion of deception towards an investigator and the resulting lack of trust should be restricted to studies that actually use deception. Such factors that might cause suspicion could be glitches in the programming of a study, flaws in the design of a study in general or the general assumption that economists or psychologists "cannot be trusted". The result will be the same as if deception had been used if a participant is suspicious and if parts of the presentation of a study are considered to be "unconvincing" by this participant. This result will be most likely a drop of trust and an increased probability of deviating or unsystematic response patterns, regardless of the actual use of deception. Taking this argument to the next level, researchers demanding a ban of deception in experimental research should demand a ban of studies that are flawed with one of these factors that might lead to a similar drop in participants trust and the studies credibility as a study incorporating deception. It can be argued in a similar vein that not only deception affects results via a pollution of the subject pool. Methodological strategies conceived in order to avoid deception might have as well a strong impact on results (Cook & Yamagishi, 2008). Supporters of the justified use of deception argue not only against the functional but also against the moral concerns of the opponents of deception. Christensen criticizes that the empirical evidence for the disapproval of the use of deception is usually based on the testimonials of single participants (Christensen, 1988), and that hardly any serious empirical research on the effects of deception has been carried out by those who

criticize its use (Smith & Richardson, 1983). The author furthermore argues in a brief summary of some empirical evidence, that most kinds of deception have no negative, but rather positive effects on participants´ well-being. What is called deception by researchers might not even be perceived as deception by the participants but rather be considered as a necessary part of the research. These conclusions are however limited by more extreme examples of deception such as Milgram´s studies on obedience (Baumrind, 1985; Kelman, 1967). Baumeister and colleagues report in keeping with the findings of Christensen that the debriefing after a rather serious deception does not necessarily have to bring about a detrimental effect. Participants performed especially well on a recall test that came after the debriefing. The knowledge about having been deceived obviously did not lead to any drop in motivation with regard to the experimental tasks or general disgruntlement (Baumeister, Twenge & Nuss, 2002: 2nd experiment).

3.3.1.4 The Use of Deception in the Presented Experimental Paradigm

The APA identifies in a pragmatic approach three conditions that legitimize the use of deception: (1) The investigator has to decide that the use of deception is justified and that alternative procedures are not feasible. (2) Participants must not be deceived regarding significant aspects of the study that would affect their willingness to participate such as physical risks and discomfort. (3) Any deception must be explained to participants preferably at the conclusion of their participation in a study but no later than the conclusion of the research.

The behavior of the five co-players in the Public Good Game constitutes a critical aspect in the present research. The aim of this goal-deviation manipulation is that contribution decisions by the IG are perceived as in line with the superordinate goal and contribution decisions by the OG are perceived as deviations from the superordinate goal. But why is it so important that this baseline (all IG members contribute in line with the superordinate goal, all OG members deviate) is kept constant by the use of deception? First it cannot be assumed that the relation between the number of deviating players and the degree of harm inflicted on the OG is linear regardless of group membership. The deviation of two players regardless of group membership cannot be expected to lead to the same extent of harm inflicted on the OG. The deviation of two OG members will lead for example to a disproportionate increase in harm inflicted on the OG compared to the case of two deviating IG members. Secondly the target of interest of the basic research questions is the example of an entire deviating OG and not individual deviants. The goal deviation has to be attributed to the deviants OG

membership. The higher the heterogeneity of OG members with regard to their deviation from the superordinate goal the lower the probability that this deviation from the superordinate goal is perceived as an inherent feature of their OG membership. A clear attribution of the deviation from the superordinate goal to the deviants´ group membership is therefore crucial. It is however highly unlikely that this benchmark pattern (all members of the IG contribute in line with the superordinate goal and all members of the OG deviate from this superordinate goal) will occur by chance. One of the main reasons for the use of deception in experimental studies is to produce situations of special interest that are unlikely to arise purely by chance (American Psychological Association, APA, 2001). Staging the contribution decisions of the five other players is therefore justified in the case of the present research. The participants receive information that the members of the OG defected thus deviating from the superordinate goal. The aim of this deception is to create homogeneity of group behavior in order to attain a benchmark in which group membership and goal-deviation are clearly related (both other IG members contribute in line with the superordinate goal, all three OG members deviate significantly from this goal)[8].

3.4 Goal-type Manipulation

It is the main hypothesis of the present research that the deviation of an OG from a minimal superordinate goal will lead to a higher level of harm inflicted on members of this deviant OG than the deviation from a maximal superordinate goal. The manipulation of goal-type (minimal versus maximal) of the superordinate goal constitutes therefore a central component of the experimental paradigm. The superordinate goal has been operationalized as contributing to a Public Good in order to make money. Maximal goals are characterized by a graded evaluation relative to an ideal goal. The closer an event is relative to this ideal goal, the better its evaluation. Minimal goals on the contrary are defined by a threshold that separates a positive evaluation of an event from a negative evaluation. The evaluation of an event in terms of a minimal goal is therefore an "either/ or"-evaluation. A manipulation of this goal-type structure of minimal and maximal goals can be modeled using variations of a Public Goods Game. Both minimal and maximal goal manipulation follow the same basic principle. The superordinate goal is identical in both Public Goods Games: Making money by contributing collectively to the Public Good. A slight change in the structure of the Public

[8] The exact feedback participants received about the contributions of the other alleged participants is given in the descriptions of the respective experiments.

Goods Game is causing players to perceive this group goal as a graded "the more the better"-maximal goal or an "either/or"-minimal goal (see Figure 3).

Minimal goal condition. A minimal goal is characterized by a dichotomous structure. The goal can either be fully achieved or not at all. In order to shape participants´ representations of the superordinate goal of contributing to the Public Good in such an "either/or" fashion a specific design of Public Good Game is introduced. A Step-Level Public Good[9] is characterized by a threshold (Figure 3) that has to be reached in order for the Public Good to be supplied (Rapoport, 1985, 1987; Van De Kragt, Orbell, Dawes, 1983). If a specific contribution rate is reached each player receives a payoff from the Public Good. Reaching this contribution threshold is therefore positive. However, a significantly higher average contribution rate is not more positive, because the payoff from the Public Good is in both cases the same, no matter by how much the threshold is exceeded. An average contribution rate of less than the specified contribution threshold is negative because no player receives a payoff from the Public Good. The structure of this step-level Public Goods Game resembles therefore the "either/or" evaluation structure of a minimal superordinate goal.

Maximal goal condition. A Graded Public Goods Game is characterized by a linear increase in payoff from the Public Good relative to the amount contributed by all players. The higher the contribution rate the linearly higher the payoff from the Public Good. The payoff from the Graded Public Goods Game resembles therefore the structure of a maximal "the more the better" superordinate goal.

[9] Some authors (e.g. Van De Kragt, Orbell, Dawes, 1983) use the term "Minimal Contribution Set" is used. For the sake of clarity the term "step-level" Public Goods Game will be used henceforth.

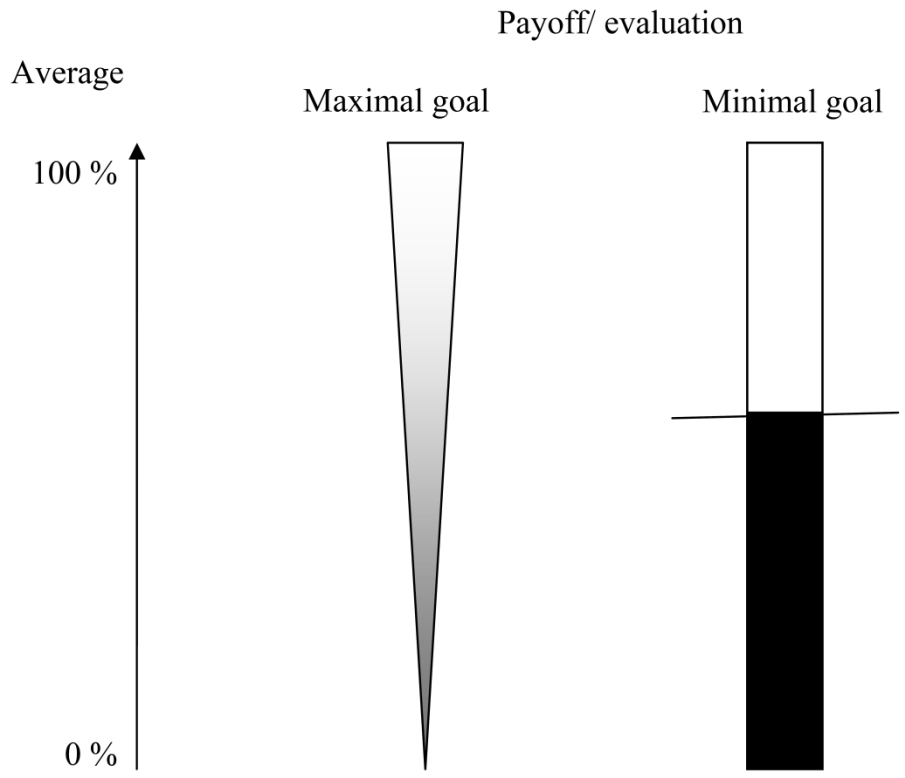

Figure 3: Step-Level and Graded Public Goods Game

3.5 Summary

A methodological paradigm has been suggested to address the basic research question of this research: Does the deviance from a minimal goal deal to a higher level of harm inflicted on members of a deviant OG than the deviance from a maximal goal? A number of methodological aspects have been implemented in this experimental paradigm. It is synergetic as different aspects of experimental games in general and the Public Goods Game in particular are used to address a multitude of these aspects: The use of experimental games most importantly ensures a *highly controlled experimental environment* while allowing at the same time the *measurement of actual negative forms of behavior* such as punishment and social exclusion. Slight modifications of a Public Goods Game can be used to manipulate participants´ *representation of a superordinate goal as either minimal or maximal*. The following chapter presents a series of experiments using this experimental paradigm in order to test the impact of a minimal versus a maximal superordinate goal on the infliction of harm on a deviant OG.

4 Experimental Studies

4.1 Study 1: Goal-type and Monetary Punishment

This first study is supposed to create a basic realization of the previously conceptualized experimental paradigm. It is supposed to incorporate a number of criteria such as a superordinate goal, an experimental manipulation of participants´ goal-type representation of this superordinate goal, behavioral measures of harm as dependent variables, and an experimentally controlled environment. It is the main hypothesis that participants with a minimal representation of the superordinate goal will be more inclined to punish members of a deviant OG monetarily than participants with a maximal representation of the superordinate goal. A secondary hypothesis assumes that this effect of goal-type (minimal/ maximal) on the monetary punishment of deviants will be mediated by negative moral emotions[10] toward the OG.

4.1.1 Methods

Design and Participants. This first experiment had one between subjects factor (goal-type: minimal versus maximal goal). Forty students from the University of Jena participated in this study. The mean age was 22.85 years, *SD* = 2.43, range: 19-34; 22 of them female and 18 male. Participants were randomly assigned either to the minimal, n =21 participants, or to the maximal, n =19 participants, goal-type condition.

Goal-type Manipulation. Goal-type was manipulated by two different types of Public Good Games. In the minimal goal condition a step-level Public Goods Game was played. In order to receive a payoff from the Public Good at least four players had to contribute their initial endowment. This threshold resembled the basic structure of a minimal goal. In the maximal goal condition the payoff from the Public Good increased gradually with the number of contributing players thus resembling the basic structure of a maximal goal.

Procedure. Participants were categorized as members of one of two minimal groups in a faked pretest (“beta-personality-type”). In the main study participants were welcomed and taken to individual sound-proof booths. No communication between participants was possible during the experimental session. The entire study took place via personal computer.

[10] „Negative moral emotions toward the OG“ will be referred to as „negative emotions“ in the following.

Participants were told that their computer was connected to the computers of five other participants participating in the same session of the study. The basic rules of a Public Goods Game based on dichotomous contribution decisions were explained (minimal goal condition: Step-level Public Goods Game; maximal goal condition: Graded Public Goods Game).

Trial round. Participants had subsequently an opportunity to compare the outcomes of different contribution-constellations in a test trial Participants were then asked a short standardized question regarding the game rules to assure their understanding of the rules of the Public Good Game.

Main round. Participants were asked in a next step to make their contribution decision in the actual Public Goods Game. Participants were informed about the contribution decisions of their co-players and the deviation of the OG players from the superordinate group goal of contributing the initial endowment[11]. This goal deviation consisted of all three players defecting. The participants´ two other IG members contributed their initial endowment to the Public Good. These contribution decisions were staged. Participants were asked to answer a list of dependent variables immediately after the notification of the OG goal deviation. Participants were paid[12], thanked and debriefed at the end of the study.

4.1.2 Dependent Measures

Monetary Punishment. Monetary punishment was measured on a seven-point scale[13] asking for participants´ inclination to deduct a share of payoff of the members of the OG ("A share of the payoff of the members with Gamma-personality should be deducted because of their decisions"). It was clear that there was no benefactor of this deduction and that it could only be interpreted as a punishment.

Negative Emotions. Nine items measured the three distinct negative emotions anger, ("I am angry at the OG", "I am annoyed by the OG", "I am enraged by the OG", Cronbach´s α = .83), contempt ("I disdain the OG", "I condemn the OG", "I am outraged by the OG", Cronbach´s α = .87), and disgust ("The OG is disgusting me", "The OG is sickening me", "I find the OG revolting", Cronbach´s α = .92).

[11] The „superordinate goal of contributing the initial endowment" will be referred to in the following as „superordinate goal".

[12] The initial endowment participants kept and the payoff from the PG summed up to 2.99 € regardless of individual contribution rate and goal-type condition. This was achieved by keeping the exact amount of the initial endowment uncertain.

[13] All seven-point scales used in the present research range from "don´t agree at all = 1" to "strongly agree = 7".

Ingroup Identification. Participants´ identification with their IG was measured with four items: "I identify with my group", "I consider myself a member of my group", "I am glad to be a member of my group", "I belong to my group". The scale ranged from "don´t agree at all = 1" to "strongly agree = 7". Identification with the IG was measured twice, once directly before the feedback about the OG deviation from the superordinate goal and once after this feedback and the other dependent variables. Identity (t_1): Cronbach´s α = .86; Identity (t_2): Cronbach´s α = .91. Participants´ identification with their IG, t_1: M = 4.90; SD = 1.49, t_2: M = 5.06; SD = 1.39, was at both measuring points significantly higher than the midpoint of the scale, t_1: t (39) = 3,81, p < .001; t_2: t (39) = 4,80, p < .001. There was a main effect of goal-type on the level of identification with the IG in t_1, minimal goal: M = 5.32; SD = 1.33; maximal goal: M = 4.43; SD = 1.56, F (1, 39) = 3.771, p = .030, η^2 = .09, as well as in t_2, minimal goal M = 5.42; SD = 1.25; maximal goal: M = 4.66; SD = 1.57, F (1, 39) = 3.126, p = .042, η^2 = .08.

4.1.3 Results

Goal-type and Infliction of Harm. An analysis of variance (ANOVA) with the independent between subjects factor of goal-type and the dependent variable of monetary punishment of the members of the OG yielded a significant main effect, F (1, 39) = 3.74, p = .031, η^2 = .09. Participants in the minimal goal condition were more likely to punish the OG more severely, minimal goal: M = 4.48; SD = 1.86, than participants in the maximal goal condition, M = 3.21; SD = 2.28.

Goal-Type and Negative Emotions. An analysis of variance (ANOVA) with the independent between subjects factor of goal-type and negative emotions toward the OG as dependent variable was conducted but no significant effect was found. Instead identification with the IG mediated the relationship between goal-type and monetary punishment of the OG. Goal-type had a significant relation with the level of identification in t_2, β = -.28, p = .043, and with monetary punishment of OG members, β = -.30, p = .032, and the level of identification with the IG in t_2 was significantly related to punishment of members of the OG independently of goal-type, β = .420, p = .007. When controlling for the level of identification with the IG in t_2, goal-type had no significant effect on monetary punishment of OG-member, β = -.199, p = .10[14].

[14] Identification with the IG in t_1 however did not mediate the relation between goal-type and punishment of the OG.

4.1.4 Discussion

This first study confirmed the *goal-type hypothesis*. Participants in the minimal goal-condition were more likely to punish members of the deviant OG than participants in the maximal goal condition. The *negative emotions hypothesis* however was not confirmed. Participants in the minimal goal condition did not report more negative emotions towards the OG than participants in the maximal goal condition. Instead the main effect of goal-type on monetary punishment of the OG was mediated by participants´ identification with their IG at t_2 after the Public Good Game. There was no mediation of the effect of goal-type on the punishment of the OG by participants´ identification with their IG at t_1. This process was unexpected. A replication of the effect of goal-type on the level of monetary punishment that was found in this first study appears crucial. A second study should clarify if the mediation of the effect of goal-type on the monetary punishment of the OG by IG identification is a systematic process or whether it should be rather attributed to the small sample size and a chance effect. Study 1 was in addition a first attempt to design and test a first design of the aspired experimental paradigm. Some general methodological issues need to be discussed in order to improve the overall experimental paradigm.

Goal-type manipulation. An important point of criticism aims at the manipulation of goal-type in this first study. The goal-type manipulation was based on a dichotomous contribution decision. Participants could either contribute their entire initial endowment or nothing at all. The dichotomous structure of this contribution decision has an intuitive fit with the idea of a minimal goal, a goal that can be fully achieved or not at all. But there is no such fit in case of a maximal goal representation. On the contrary, the contribution option allows individual players to act absolutely in favor of the superordinate goal by contributing the entire initial endowment or nothing at all. This might - at least in the maximal goal condition - constitute a goal-type manipulation on an individual level that conflicts with the goal-type manipulation at the group level. The manipulation of the superordinate goal as maximal might have been flawed by the dichotomous contribution structure at the individual level.

Goal-setting. There were no clues for participants indicating that the superordinate goal of contributing to the Public Good in order to make money was a shared one. The experimental setup provided no information about the co-players except for their group-membership, no face-2-face interaction, and no option to communicate. Participants were thus confronted with a great deal of uncertainty regarding the behavior of the other players. However, participants´ trust that their co-players will contribute to the Public Good is essential for their own

contribution to the Public Good. The introduction of the group goal of making money by contributing to the Public Good should therefore be greatly improved if participants had more cues that this group goal is a shared one among group members.

Deception and credibility. The experimental implementation of this first study incorporated a certain extent of deception in order to keep the degree of OG-deviation constant across all participants. The information participants received about the contribution decisions of their co-players were staged. Anecdotal reports by participants suggest that some of the participants doubted the authenticity of the interactions with the other players. Any doubts regarding the authenticity of participants´ interactions in this study are potentially detrimental to the results of a study as they may lead to unexpected patterns in terms of the dependent variables. Therefore an improved overall credibility of the study has to be another concern of a replication of this first study.

Summary. Some conceptual concerns suggest that the goal-type manipulation of the superordinate goal used in this first study could be enhanced by avoiding a dichotomous contribution format for the Public Goods Game that might interfere with the goal-type manipulation. It has been argued that the abstract and minimal design of the study might have resulted in a suboptimal introduction of the superordinate goal of making money by contributing to the Public Good. Some cues that the goal of making money by contributing to the Public Good is a shared one among group members should reduce participants´ uncertainty and should improve the introduction of the superordinate goal. A second study should thus replicate the effect of goal-type on monetary punishment of members of a deviant OG that was found in this first study. Secondly, it should clarify whether this effect is mediated by identification with the IG or negative emotions towards the OG. This second study should be based on an improved design of the experimental paradigm used in Study 1.

4.2 Study II: Goal-type and Monetary Punishment: Replication

The results of the initial study made clear that certain aspects of the experimental paradigm need to be improved. The most important enhancements are a more effective introduction of the superordinate goal, a more elegant goal-type manipulation, and an improvement of the overall credibility of the study because of the use of deception. It seems probable that it was not so much the occurrence of one systematic pattern that might have lead some participants doubt the authenticity of the players´ interactions in the Public Good Game of Study 1. But rather a co-occurrence of three of these patterns: The same number of participants from IG and OG were assigned to the same session (1), exactly three of six participants contributed and three defected (2) and all IG members contributed and that all OG members defected without exception (3). A first step to prevent suspicions by participants is therefore to inform participants that they were not randomly assigned to the sessions but that the pattern of three players from their IG and three players from the OG in one session was created deliberately. This way only two instead of three suspicious patterns co-occur.

Another concern regarding the experimental paradigm used in Study 1 addressed the introduction of the superordinate goal of making money by contributing to the Public Good. The next section deals therefore with the improvement of the goal setting process.

4.2.1 Goal-type Manipulation

Some general methodological concerns and unexpected empirical findings in the first study call for an improved goal-type manipulation. One major point of criticism was the dichotomous contribution option that had rather a fit with the dichotomous structure of a minimal goal that was possibly obstructing a manipulation of the representation of the superordinate goal as maximal. A slightly different combination of contribution option and goal-type manipulation is therefore suggested for this second study. This approach utilizes a graded contribution option. As in the previous study a step-level Public Good Game is used in order to create a minimal goal representation and a graded Public Good Game to create a maximal goal representation of the superordinate goal. The main difference is that the contribution decision is not dichotomous but graded. Participants have the option to contribute between 0 % and 100 % of their initial endowment. The group goal is accordingly represented by an average contribution rate. The maximal goal is represented by an ideal average contribution rate of 100 %. The minimal goal is represented by a threshold of the average contribution rate of a certain percentage. This rate has to be exceeded in order to receive a

payoff from the Public Good. The threshold in the minimal goal condition in this study was set at an average contribution level of 60 % of the initial endowment. The basic approach of manipulating participants´ goal-type representation with slightly different designs of a Public Good Game is retained in principle. But participants in the maximal goal condition are no longer forced to make their contribution decision in a dichotomous "either/ or"-format that is rather fitting the structure of a minimal goal. The use of a graded contribution option should therefore result in a more coherent goal-type manipulation. A second positive effect of the graded contribution option is an increased overall credibility of the study. Feedback from participants of the first study suggested that it was the co-occurrence of certain patterns that raised their suspicions regarding the study and participants interaction in it. The fact that all IG-members contributed to the Public Good in the first study and all OG-members did not contribute constitutes such a pattern. The introduction of a graded contribution option allows a degree of variance within the groups. This variance should improve the overall credibility of the study.

4.2.2 Introduction of a Superordinate Goal

A third concern addresses the improvement of the introduction of the superordinate goal of collectively contributing to the Public Good. A broad variety of factors have been identified that have the capability of increasing participants´ probability to cooperate in Social Dilemmas. In his work on strategic solutions to social dilemmas Axelrod (1984) identified three basic requirements for cooperation to occur: 1. Individuals have to be involved in an ongoing relationship. 2. Individuals must be able to identify each other. 3. Individuals must have information how the other participants have behaved in the past. Participants had in the initial study of the present research no information if the goal of making money via contributing to the Public Good was a shared group goal until they received the feedback about the contributions of their co-players only seconds before the dependent variables. In these few seconds participants had to process the implications of the contribution decisions for the Public Good, the implications for their own payoff, they had to understand that therse contribution decisions were in line with the group identities, attribute these differences to the respective group membership and respond accordingly. This might have exceeded cognitive capacities or willingness of some of the participants which in turn might have had a detrimental effect on their representation of contributing to the Public Good as a superordinate goal. Participants might need a little more information and time to process this information in order to achieve a less ambiguous, more reliable introduction of the superordinate goal. The

introduction of the superordinate goal of contributing to the Public Good and the actual deviation from this goal by the OG should therefore be separated.

The superordinate goal of collectively making money via contributing to the Public Good should be introduced in the first round of the Public Goods Game. In this first round only members of the IG play, contributing a considerable share of their initial endowment[15]. This first round serves the purpose to inform participants with a minimal amount of information that the other players in their own group followed the same contribution goal and do what they say. Participants are therefore - referring to Axelrod´s stated conditions for cooperation to occur - 1. involved in an ongoing relationship 2. with the ability to identify the other players individually with a letter-code and more importantly as members of their IG and 3. receive information that these members of their IG contributed to the Public Good. This should decrease the level of uncertainty regarding the superordinate goal, support participants´ perception of contributing to the Public Good as a shared superordinate goal and increase contribution decisions in line with this goal.

4.2.3 Methods

Design and Participants. This second study had one between subjects factor (goal-type: minimal versus maximal goal). Sixty students from the University of Jena participated in this study. The mean age was 22.08 years (*SD* = 3.51, range: 18-35); 42 of them female and 18 male. Participants were randomly assigned to a minimal (30 participants) respectively maximal (30 participants) goal-type condition.

Procedure. The basic experimental setup was the same as in the first study. The following description of the procedure used in this second study addresses therefore only aspects that changed compared to the first study.

Introduction Superordinate Goal. Participants were asked in an initial trial round to contribute an amount between 0 % and 100 % of their initial endowment. Participants knew neither the exact amount of the initial endowment nor the exact number of rounds played. Participants were notified a few seconds after their own contribution decision that the other two members of their IG had contributed 70 % and 90 %. They were then informed about the average contribution rate of all three players, the amount of money kept and the payoff from the common account. By varying the exact amount of the initial endowment payoffs for all

[15] The exact level of the contributions reported to the participants is given in the descriptions of each of the respective experiments.

participants could be kept at exactly the same level[16]. Right after the end of the first round the second round started. It was played by all six members of the SOG. Shortly after their own contribution decision participants were notified about the contributions of their five co-players. The information about the contribution levels of their co-players was exactly the same for all participants, across goal-type conditions. Participants were notified that the two other members of their IG contributed 60 % respectively 80 % of their initial endowment, thus keeping in line with the group goal. The members of the OG contributed only a relatively small amount of their initial endowment (10%, 30% and 40 %) thus committing a goal-deviation[17]. The average contribution rate was calculated for both IG and OG and displayed. Participants were asked afterward to answer a set of dependent variables. After the study participants were paid, thanked and debriefed. *Contribution Decisions*. Second round contributions were similar in the second round that was allegedly played by participants from both IG and OG. Participants in the minimal goal condition contributed on average 71 % *SD* = 14.37, of their initial endowment, participants in the maximal goal condition 76 %, *SD* = 21.61.

4.2.4 Dependent Measures

Monetary Punishment. Monetary punishment was measured on a seven-point scale asking for participants´ support of a partial deduction of payoff for the members of the OG in the same way as in the previous study.

Negative Emotions. A combined scale of negative emotions towards the OG comprised of three items was designed. The items were targeted directly at the OG ("I am mad at the Gamma-group", "I am outraged by the Gamma-group", "I think the Gamma-group is disgusting").

[16] The initial endowment participants kept and the payoff from the PG summed up to 2,99 € regardless of individual contribution rate and goal-type condition. This was achieved by keeping the exact amount of the initial endowment uncertain.

[17] All participants received in the initial trial round a payoff of 3,00 Euro (kept share of the initial endowment + payoff from the Public Good), regardless of goal-type condition and their individual contribution decision. This was achieved by varying participants´ initial endowment relative to their contribution decision. In the second round two variables had to be adjusted: Participants total payoff from the main round (kept share of the initial endowment + payoff from the Public Good) and the amount of money participants missed out on because of the OG´s deviation of the superordinate goal. Participants´ total payoffs in the main round varied between 2,88 Euro and 3,45 Euro in the maximal goal condition and between 2,66 Euro and 3,00 Euro in the minimal goal condition.

Ingroup Identification. Participants´ identification with their IG was measured before the Public Goods Game on 4 items: "I identify with my group", "I consider myself a member of my group", "I am glad to be a member of my group", "I belong to my group"[18]. Participants´ identification with their IG, $M = 5.25$; $SD = 1.40$, was significantly higher than the centre of the scale, $t\ (59) = 6{,}913$, $p < .001$. There was no main effect of goal-type on level of identification, neither at t_1, $\beta = -.05$, $p = .717$, nor a t_2, $\beta = -.02$, $p = .891$.

4.2.5 Results

Goal-type and Infliction of Harm. An analysis of variance (ANOVA) with the independent between subjects factor of goal-type condition yielded a significant effect on punishment of OG-members by deducting money from them, $F\ (1, 59) = 5.61$, $p = .011$, $\eta^2 = .09$. Participants in the minimal goal condition punished the OG significantly stronger than participants in the maximal goal condition, minimal goal*:* $M = 4.53$; $SD = 2.06$; maximal goal: $M = 3.30$; $SD = 1.97$.

Goal-Type and Negative Emotions. Participants in the minimal goal condition reported significantly more negative emotions with regard to the OG than participants in the maximal goal condition, $F\ (1, 59) = 4.67$, $p = .018$, $\eta^2 = .08$, minimal goal: $M = 3.77$; $SD = 1.53$; maximal goal: $M = 2.92$; $SD = 1.50$. In order to test the indirect effect bootstrapping was used (N=2000) to determine the standard error (Preacher & Hayes, 2004; 2008). The indirect effect of goal-type on the dependent variable monetary punishment was mediated through negative emotions towards the OG, BCa (95%) = [-1.1749, -.0330].

4.2.6 Discussion

The main hypothesis could be confirmed in this second study, replicating the finding of Study 1: Participants in the minimal goal condition punished members of the OG more severely for their deviation from the superordinate goal than participants in the maximal goal condition. This process was not mediated by the level of identification in t_2 as in the first study, but by negative moral emotions towards the OG. This difference in results might be attributable to the remodelled goal-type manipulation. This more precise manipulation of goal-type and the improved overall design of this second study might have helped to reveal negative emotions as the mediator of the effect of goal-type on harm inflicted on the OG as it was initially hypothesized.

[18] All analyses with the independent variable goal-type are one-tailed analyses in this chapter.

There was nevertheless a minor but possibly important difference in the goal-setting between minimal and maximal goal condition. The ideal goal in the maximal goal condition was set by logic. The goal of making money by contributing to the Public Good as a group suggests that the ideal average contribution rate is 100 % of the initial endowments. In the minimal goal condition in contrast a threshold of 60 % of average contributions was given by rule as superordinate goal. The differences in payoff between the step-level Public Good Game and the graded Public Good Game were minimized by keeping the exact amount of the initial endowment unknown. The remaining differences did not affect the results of this study. But some psychological differences remain nevertheless. The goal in the maximal goal condition was implied indirectly by logic. The superordinate goal in the minimal goal condition was given by game rule respectively the experimenter. It could be argued that the additional factor of authority increased participants´ tendency to punish deviants from this goal. This could have been the case either because a deviation from a superordinate goal given by rule and external authority appeared more severe or simply because the punishment of such a transgression was easier to justify. But the precise process of a possible impact of the factor "source of superordinate goal" (logic versus rule) on the extent of punishment of deviants is only secondary. It is more important to note that the factor "source of superordinate goal" is a possible alternative explanation for the effect of goal-type on the severity of monetary punishment found in this second study. Future studies will have to control experimentally for the source of the superordinate goal as a potential confound of goal-type.

A number of implications for a follow-up study have been identified. The effect of goal-type on monetary punishment has been shown and replicated in the first two studies. It seems therefore advisable to test if the scope of this effect of goal-type on punishment of the OG can be extended to other forms of harm inflicted on members of a deviant OG. In addition another replication of the effect of goal-type on the infliction of harm on a deviant OG appears necessary. In the first study identification with the IG mediated the effect of goal-type on monetary punishment. In the second study no such mediation could be found. The effect of goal-type on monetary punishment was instead mediated by negative emotions as initially hypothesized. Another replication might help to clarify the relation between goal-type and the infliction of harm. This replication should control in addition the "source of superordinate goal" as a potential alternative explanation for the effect of goal-type on the infliction of harm on the deviant OG.

4.3 Study III: Social Exclusion

4.3.1 Introduction

This third study is set to address multiple aims. First of all the source of the superordinate goal has to be controlled for as a possible confound of the factor goal-type manipulation. This addresses a major methodological concern of the second study. A new goal-selection mechanism is therefore presented. It is over and above another aim of this study to exclude differences in the payoffs in minimal and maximal goal condition as possible alternative explanations for the effect of goal-type on the infliction of harm. This third study should help last but not least to clarify the processes of the effect of goal-type on the infliction of harm on members of a deviant OG: Is it mediated by identification with the IG - as the first study suggested -or by negative moral emotions towards the OG -as the second study implied? This third study is in addition supposed to test if the impact of goal-type can be extended to other forms of harm than monetary punishment. The exclusion from a superordinate group constitutes such an alternative form of harm. This study will therefore introduce the social exclusion of members of the OG that are present in a session (social exclusion) as well as the social exclusion of members of the OG, that are not present in a session (generalized social exclusion) as new dependent variables.

4.3.2 Social Exclusion

Various forms of harm that might be inflicted on members of a deviant OG in an experimental context have already been mentioned. These forms of harm - such as unpleasant noise (Mummendey et al, 1992) or monetary fines (Fehr & Gächter, 2004) involve a direct cost against the will of their target. The following chapter presents another archetypical form of harm: Social exclusion. Social exclusion might be defined as the exclusion of individuals from a social group against their will. One of its crucial characteristics is its capacity to facilitate harmful treatments of others as it removes them from the normative and moral protection of their group membership. The following chapter will therefore review some of the literature on the phenomenon of social exclusion both from the perspective of the excluded as well as the excluding.

4.3.2.1 The Phenomenon of Social Exclusion: Motivation for Inclusion

Being excluded from a social group is usually seen as an extremely negative event that is often associated with painful emotional experiences. The fact that being excluded has such a

negative valence raises the question why inclusion in a specific social group is so highly estimated. It has been argued on many occasions and for many social species that the inclusion in groups provides individuals with resources that prove advantageous with regard to survival, mating and raising offspring. Protection against the forces of nature such as predators and rivaling groups but also the division of labor are just some examples of the positive effects of groups. The lack of social inclusion on the flipside might as well imply severe consequences as some examples of research on social exclusion among primates might illustrate. Infants of highly socially integrated baboons have been shown to be more likely to grow older than age one than infants of less socially integrated mothers (Silk, Alberts, & Altmann, 2003). Kling and colleagues showed furthermore that vervets and rhesus monkeys who showed a low interest in social contact as a result from brain site lesions were often excluded after re-release to the wild and died without the protection of their conspecifics (Kling, Lancaster, & Benitone, 1970). This research supports the general idea that a strongly integrated social live increases the chances of surviving, reproducing, and raising offspring to a reproductive age among social animals. The observation that humans as one kind of social animal benefit from social inclusion in a broad sense via discriminative cooperation has found support from a variety of theoretical perspectives. In contrast to these approaches the same argument has been made from less ”gene-centric” approaches from a rather socio-biological or economic, individual-focussed background. These approaches argue that sociality has not developed as an afterthought, an update to individual psychological processes but rather that the inclusion in groups constitutes a crucial part of the environment the human psyche evolved in. Cognition and sociality are from this perspective inextricably intertwined (Caporael, 1997). The observation that people have a powerful need to belong may be seen, following this line of reasoning, as the result of an adaptive process that was shaped by the importance of sociality and groups as crucial factors in the environment (e.g. Baumeister & Leary, 1995; Maslow, 1968). Social exclusion constitutes a fundamental threat to this need. However there are two sides to the phenomenon of social exclusion, the excluded and the excluding.

4.3.2.2 Being Excluded: Consequences of Social Exclusion

Ample research has highlighted the potentially detrimental effects of the experience of being excluded from a social group. The following short review of research on the effects of being excluded might illustrate the scope and severity of the experience of being excluded. Along the many negative effects of the experience of being excluded from a social group is an

increased probability to take irrational, self defeating risks, choose unhealthy over healthy behaviors and show a greater tendency to procrastinate (Twenge, Catanese & Baumeister, 2002). Stillman and colleagues showed in a series of studies, that social exclusion affects the global perception of life as meaningful in a negative way (Stillman, Baumeister, Lambert, Crescioni, DeWall, Fincham, 2009). The prospect of social exclusion might even affect basic cognitive capacities. Participants who were told they would end up alone in live showed a significantly decreased performance in complex cognitive tasks such as logic and reasoning (Baumeister, Twenge & Nuss, 2002; see also: Twenge, Catanese & Baumeister, 2003) as well as decrements in the effort of self-regulation (Baumeister, DeWall, Ciarocco & Twenge, 2005). This effect was specific to social exclusion and not mediated by mood. Similarly Twenge and colleagues showed that the expectation of ending up alone later in life had a negative effect on feelings of empathy towards a person in need of help, leading to a reduction in prosocial behavior. This result was replicated in a mixed-motive game where being socially excluded led to a lower level in empathy with the co-player and a significantly lower level in cooperation (Twenge, Baumeister, DeWall, Ciarocco, & Bartels, 2007).

Socially excluded participants reported even under superficial conditions such as a simple ball tossing game lower levels of belonging, self esteem, control perception and meaningful existence (Williams, Cheung & Choi, 2000). These results could be replicated in a computer based version of the ball throwing game, the "cyberball"-paradigm (Williams & Jarvis, 2006). Comprehensive research highlights ostracism[19] (Williams, 1997) as a factor increasing aggression towards others (DeWall, Twenge, Gitter, & Baumeister, 2009; Twenge, Baumeister, Tice, & Stucke, 2001). Chow and colleagues pointed out the importance of specific anger in response to the exclusion as a mediator for antisocial behavior (Chow, Tiedens & Govan, 2007). Williams and co-authors showed that the relation between the experience of being ostracized and aggression is moderated by threats to individual control needs (Warburton, Williams & Cairns, 2006). The perception of social exclusion as painful seems to be so stable that not even mitigating circumstances such as a cost for inclusion and retaining money for social exclusion change this relation (van Beest & Williams, 2005). Ostracism might even lead to emotional distress when it is inflicted by a disdained OG (Gonsalkorale & Williams, 2007). The importance of inclusion in social groups for humans has been described earlier. It might be argued similarly that threats to the social inclusion and

[19] Ostracism may be defined as being ignored and often occurs without explanation or the explicit infliction of harm. However, it has been pointed out that virtually no empirical research points to different implications of ostracism and social exclusion (Williams, 2007). Both terms are accordingly often used interchangeably.

threats to physical wellbeing share respectively the resulting social and physical perceptions of pain. A connection of this highly negative affect to social exclusion might be interpreted from an evolutionary perspective in a similar way as the main function of physical pain. Physical pain serves as a cue to react to a threat to an individual´s physical integrity (e.g. the pain resulting from a thorn about to pierce the skin). Similarly social pain may be interpreted as a cue reminding the individual of the negative implications of being socially excluded (e.g. Krebs, Stephens, & Sutherland, 1983). Herman and Panksepp (1978) suggested that neurological processes for separation distress would represent an evolutionary elaboration of an endorphin-based pain network (Hermann & Panksepp, 1978; Panksepp, 1998). The perception of social pain affect fulfills according to MacDonald and Leary two important functions in terms of social regulation needs. First it enables a learning process to take place in order to avoid inclusion-threatening situations and minimize the probability of actual exclusion. Secondly a close connection of pain and threat-defense response mechanisms should enable a quick link between social exclusion cues (Kerr & Levine, 2008) and defensive reactions to regulate the individual´s social inclusion status (MacDonald & Leary, 2005). A neuroimaging study conducted by Eisenberger and colleagues supports this assumption that the neurological bases of social pain are similar to those of physical pain (Eisenberger, Lieberman, Williams, 2003). Further research showed that the baseline sensitivity to physical pain predicts the sensitivity to social rejection and that the heightened social distress increases the sensitivity to physical pain (Eisenberger, Jarcho, Lieberman & Naliboff, 2006), thus gaining more evidence in support of the assumption that pain distress and social distress share neurological substrates. Zhong and Leonardelli could show in line with theories on embodied cognition that the experience of social exclusion is not only proverbially but literally “cold”. Participants in an experiment recalling a social exclusion experience gave lower estimates of room temperature and had a greater desire for warm food and drink than participants that were included in a social group (Zhong & Leonardelli, 2008).

4.3.2.3 Excluding Others: Social Exclusion as a Pathway to the Infliction of Harm on Others

But what about the perspective of the excluding group members (for an overview: Fritsche & Schubert 2009)? Many reasons incentivize social exclusion from a functional perspective. A first - although somewhat trivial - point is based on the logical observation that there is no inclusion without exclusion. “In order for sociality to be functional, there must be “brakes” on sociality” (Kurzban & Leary, 2001). Further reasons for social exclusion to appear are 1. from an evolutionary perspective the avoidance of diseases, 2. the fact that the excluded individuals

have proven to be unreliable cooperation partners in important group endeavors or simply 3. that a higher gain can be achieved for the remaining group members by exploiting the excluded group members (Kurzban & Leary, 2001). All reasons for social exclusion however boil from a functional perspective down to the simple tradeoff between the exclusion of dysfunctional group members and the possible costs accompanying the exclusion of group members in question and the costs of keeping them included. The factors influencing this tradeoff are often not as clear-cut as for example in the case of incapacitation via imprisonment of criminal offenders (e.g. Zimring & Hawkins, 1995; Darley, Carlsmith, & Robinson, 2000). The functionality of social exclusion should therefore be narrowly defined as functional in relation to a specific group goal. Social exclusion serves a function equivalent to a punishment option if a group goal is conceptualized as a Public Goods Game: Being excluded represents a possible cost for defectors. The mere threat of being excluded from a group shifts the cost/benefit tradeoff for defection, thus rendering defection more expensive and therefore less attractive. An option to exclude defectors respectively low cooperators in a Public Goods Game framework should therefore result in higher levels of cooperation for the remaining group members. Potential processes for this effect are the incapacitation of excluded defectors and a shift in incentives to cooperate for the remaining group members (Ellsworth & Ross, 1983; Vidmar, 1974, Vidmar & Ellsworth, 1974).

Social exclusion can also facilitate the infliction of harm on others such as torture, killing or even genocide that would be unthinkable and unjustifiable for members of a shared IG (Bandura, Underwood, & Fromson, 1975; Staub, 1989). Moral exclusion constitutes a specific form of social exclusion that has been defined as the exclusion of individuals or groups from one´s scope of justice (Opotow, 1990, 1995). The inclusion in a group implies the application of certain moral standards to its members. Violations of these standards have to be justified. The moral exclusion of individuals and groups from this protective shield of a group membership enables the legitimate infliction of harm on any excluded individuals of groups. This may imply in extreme cases a categorization of deviants into extremely negative social categories which are excluded from the ultimate group of humanity thus removing its members from the moral protection affiliated with being human (Bar-Tal, 1989; Haslam, 2006; for a perspective of the target: Brock & Haslam, 2010). Because of this significant importance of social exclusion as a specific form of harm, two measures of social exclusion are introduced in this third study: the social exclusion of members of the OG that are present in a session (social exclusion) as well as the social exclusion of members of the OG, that are not present in a session (generalized social exclusion).

This third study features in addition a methodological improvement. A goal-setting process is introduced in order to keep the source of the superordinate goal constant across minimal and maximal goal condition, thus addressing a major methodological concern of the second study.

Goal-type manipulation and Group-goal setting. The goal-type manipulation was identical to the one used in the second study. In this study no explicit superordinate goal the 60 % threshold in the minimal goal condition of the second study was given. Instead a goal setting mechanism was introduced. Participants were told that the group could set a strictly nonbinding, informal goal, for an average contribution level. One participant was allegedly picked at random by a computer program in order to propose such a superordinate goal represented by an average contribution level between 10 % and 100 %. The other two participants could either accept or reject this contribution goal. It was not made explicitly clear if the actual scope of this informal goal was only the IG or if it was communicated in the entire SOG (IG and OG). Every participant was "chosen" as the player to propose this group-goal. Participants received subsequently a short feedback, that both IG co-players had accepted the proposed contribution goal, regardless of the exact level of this goal. This selected goal functioned in the minimal goal-type condition as the threshold that had to be reached in order to receive a payoff from the Public Good. *Contribution Decisions*. Participants in the minimal goal condition contributed on average 62 % of their initial endowment, *SD* = 22.98, and 75 % in the maximal goal condition, *SD* = 17.11.

Goal Deviation: The selected superordinate goal was supported by the contribution decisions of the two IG-players who contributed at least at the level of the superordinate goal. The goal-deviation of the OG was adapted to the selected superordinate goal. The perceived magnitude of harm has been shown to have a decisive impact on the extent of responding punishment (Carlsmith et al., 2002). The relative goal-deviation (magnitude of harm) was therefore controlled for. The entire OG contributed always half of the superordinate goal (e.g. if the selected superordinate goal was 80 % the three players of the OG contributed on average 40 %). This proportionate relation of superordinate goal and average contribution level of the OG was chosen instead of a linear relation[20]. The relative deviation of the OG was

[20] A linear relation would have meant that the absolute difference between selected superordinate goal and average OG-contribution had been kept constant. The proportionate relation seems more appropriate in order to keep the various levels of eligible superordinate goals comparable. This way the relative magnitude of harm inflicted by the OG can be kept constant across all eligible goal levels. *Example:* If an average contribution level of 80 % is selected as superordinate goal, the OG contributes on average 40 %, thus committing a goal deviation of contributing on average only half of the aspired superordinate goal. If an average contribution level of 40 %

therefore constant not only across goal-type conditions but also across the different levels of the selected superordinate goal.

Goal-type Manipulation. The same goal-type manipulation was used as in the second study. Three items assessed on a seven-point scale participants´ representation of the goal-type, e.g. "The average contribution level of […] must be achieved"; "The average contribution level of […] should ideally be achieved"; Cronbachs α = .67. The items were adjusted to the contribution level participants had chosen previously. Participants in the minimal goal condition scored significantly higher on the goal-type manipulation check, F (1, 46) = 6.410, p = .015, η^2 = .12, indicating a stronger minimal goal representation than participants in the maximal goal condition, minimal goal: M = 5.28; SD = 1.16; maximal goal: M = 4.35; SD = 1.34.

4.3.3 Methods

Design Participants. This third experiment had one between subjects factor (goal-type: minimal versus maximal goal). Forty-seven students from the University of Jena participated in this study. Thirty-five participants were female, 12 male. The mean age was 21.87 years, SD = 1.92, range: 18-27**.**

Procedure. The basic experimental setup was the same as in the second study. The following description of the procedure used in this third study addresses therefore only aspects that changed compared to the second study.

Trial Round. The superordinate goal was set as described above. Players were informed that the first round was played only by members of their own IG. Participants were asked to contribute an amount of their initial endowment between 0 % and 100 %. Participants knew neither the exact amount of the initial endowment nor the exact number of rounds played. Participants were informed about the contribution decisions of their IG-members a few seconds after their own contribution decision. These contribution decisions were related to the chosen group goal-level. One of the two IG-members always contributed at the goal-level (e.g. 40 % if a superordinate goal level of 40 % had been selected) and the other IG-member contributed 10 % above the superordinate goal level (e.g. 50 % if a superordinate goal level of 40 % had been selected). Participants were then informed about the average contribution rate

was selected as superordinate goal, the OG contributes on average 20 %, thus committing a goal deviation of contributing only half of the aspired superordinate goal. The relative goal deviation is the same regardless of the level of the selected superordinate goal.

of all three players and the payoff from the Public Good. All participants received - regardless of their own contribution-decisions - 4 Euros from the first round.

Main Round. The second round with all six players started right after the first round. Participants were notified about the contributions of their five co-players shortly after their own contribution decision (0 % - 100 % of their initial endowment). The information about the contribution levels of their co-players was exactly the same across both goal-type conditions. Participants were notified that the two other members of their IG had contributed at the selected superordinate goal level respectively 10 % above it, thus keeping in line with the superordinate goal. The members of the OG contributed only half of the selected group goal, thus committing a goal-deviation. The average rate of contribution was calculated for each subgroup and displayed. Participants in both conditions received 2.50 € in the second round[21]. All participants received therefore exactly the same feedback about their payoffs regardless of goal-type condition. All participants received a total amount of 6.50 Euro for their participation in the study.

4.3.4 Dependent Measures

Monetary Punishment. The same measure of monetary punishment of the OG was used as in the previous studies. Two measures of exclusion were added as further examples of harm inflicted on the deviant OG.

Social Exclusion. The first measure of exclusion measured on a seven-point scale if participants wanted to exclude the OG from future rounds ("The group with Gamma-personality should be excluded from future rounds because of their decision").

Generalized Social Exclusion. A second item measured the generalized social exclusion of OG-members from the SOG. This item was introduced as a separate study and was assessed after the actual computer-based study in a pen-and-paper format. Participants were told a cover-story that explained why they had to select eight participants for a future study from a data-base containing participants of former studies willing to participate in future, highly profitable studies. Although participants of the alleged earlier studies were kept anonymous on the list, their group-identity as either members of the beta- or gamma-group was evident. There was a time gap of about 10 minutes between the first exclusion item right after the

[21] The equality of payoffs in both conditions was achieved by informing participants only about the sum of the payoff for the respective round, and not the single parts of this sum (kept amount of the initial endowment and amount received from the Public Good).

information of the contribution rates of IG and OG players and the measure for generalized social exclusion. All three measures of harm were significantly correlated: Monetary punishment and exclusion from future rounds: $r = .672$, $p < .001$; monetary punishment and generalized social exclusion: $r = .349$, $p = .015$; exclusion from future rounds and generalized social exclusion: $r = 315$, $p = .029$.

Negative Emotions. The same measures as in the second study were used to measure negative moral emotions toward the OG and identification with the IG.

Ingroup Identification. Participants´ identification with their IG, $M = 5.11$; $SD = 1.28$, was significantly higher than the centre of the scale, $t(46) = 6.00$, $p < .001$. There was neither in t_1 (F (1, 46) =.06, $p = .808$, minimal goal: $M = 5.07$; $SD = 1.55$; maximal goal: $M = 5.16$; $SD = .90$.) nor in t_2 (F (1,46) = .003, $p = .959$, minimal goal: $M = 5.29$; $SD = .99$; maximal goal: $M = 5.27$; $SD = 1.29$.) a main effect of goal-type on the level of identification with the IG.

4.3.5 Results[22].

Goal-Type and Infliction of Harm Participants in the minimal goal condition were more likely to punish members of the deviating OG more severely, F (1,46) = 4.34, $p = .022$, $\eta^2 = .09$, more likely to exclude members of the OG from future rounds, F (1, 46) = 5.95, $p = .010$, $\eta^2 = .12$, and more likely to exclude other members of this group from the participation in future profitable studies, F (1, 46) = 3.41, $p = .035$, $\eta^2 = .07$, than participants in the maximal goal condition (Table 2).

Goal-Type and Negative Emotions. An analysis of variance (ANOVA) with the independent between subjects factor of goal-type condition revealed a significant effect of goal-type on the combined scale of negative emotions towards the OG, $F(1, 46) = 4.54$, $p = .020$, $\eta^2 = .09$. Participants in the minimal goal condition reported more negative emotions towards the OG than participants in the maximal goal condition (Table 2).

[22] All analyses with the independent variable goal-type are one-tailed analyses in this chapter.

Table 2: The Effect of Goal-type on the Infliction of Harm and Negative Emotions toward a Deviant OG

	Goal-type			
	Minimal Goal		*Maximal Goal*	
Form of Harm	*M*	*SD*	*M*	*SD*
Monetary Punishment	4.46	2.00	3.32	1.76
Social Exclusion	4.54	2.34	3.05	1.81
Generalized Social Exclusion	5.69	1.67	4.82	1.59
Negative Emotions	3.37	2.10	2.24	1.45

A series of regression analyses was performed to analyze whether moral emotions mediate the relation between the manipulated factor goal type (minimal versus maximal) and the infliction of harm on the OG (monetary punishment, social exclusion from future rounds and generalized social exclusion). First, goal type had a significant impact on negative moral emotions, $\beta = -.30$, $p = .020$, and on monetary punishment, $\beta = -.29$, $p = .022$. Negative moral emotions predicted the degree of punishment, $\beta = .67$, $p < .001$. When controlling for negative moral emotions, goal type had no longer a significant effect on monetary punishment, $\beta = -.10$, $p = .192$. To test the indirect effect bootstrapping (N=2000) was used to determine the standard error (Preacher & Hayes, 2004; 2008). The indirect effect of goal type on punishment via negative moral emotions was significant, BCa (95%) = [-1.4974, -.1289].

Goal type had a significant impact on the exclusion of OG members from future rounds, $\beta = -.34$, $p = .010$. Negative moral emotions predicted the exclusion of OG-members from future rounds, $\beta = .57$, $p < .001$. When controlling for negative moral emotions, goal type had no longer a significant effect on the exclusion of OG members from future rounds, $\beta = -.18$, $p = .075$. The indirect effect of goal type on the exclusion of OG members from future rounds via moral emotions was significant, BCa (95%) = [-1.4657, -.1229].

Goal type had a marginally significant effect on generalized social exclusion of OG members, $\beta = -.26$, $p = .036$. Negative moral emotions predicted the generalized social exclusion of OG-members, $\beta = .66$, $p < .001$. When controlling for negative moral emotions, goal-type had no longer a significant effect on the generalized social exclusion of OG

members, β = -.07, p = .351. The indirect effect of goal type on the generalized social exclusion of OG via moral emotions was significant, BCa (95%) = [-1.3117, -.1032].

4.3.6 Discussion

This third study was a redesign of Study 2. Its first aim was to replicate the effect of goal-type on the infliction of harm on members of the OG that was found in the first two studies while controlling experimentally for several factors that were confounded with the factor goal-type in the first two studies. In this third study the differences in payoffs participants missed out on because of the deviant OG in minimal and maximal goal condition were experimentally controlled for. Differences in payoffs from the two different types of Public Goods Game that were used to manipulate goal-type could therefore be ruled out as alternative explanations for the effect of goal-type on the infliction of harm on the OG. Secondly, the level of the superordinate goal in both minimal and maximal goal condition was selected by the participants themselves. The goal-selection was therefore identical in both minimal and maximal goal condition. The source of the superordinate goal could therefore be excluded as an alternative explanation for the main effect of goal-type on the infliction of harm on a deviating OG.

The main hypothesis could be confirmed. Participants in the minimal goal condition were more likely to punish members of the OG thus replicating the effect of goal-type on monetary punishment of the OG that was found in the first two studies. They were in addition more likely to exclude members of the OG from the participation in future profitable rounds of the Public Good Game. A measure of generalized social exclusion showed that this effect was not restricted to the members of the OG that were committing the goal deviation but applied as well to members of the OG that did not commit any deviation, that were anonymous and not known by the participants. This effect is even more notable keeping in mind that the group identities were minimal and content-free. The effects of goal-type on various measures of harm inflicted on the OG were mediated by negative moral emotions towards the OG as in the second study. There was no main effect of goal-type on the level of identification, rendering the alternative process of a mediation of the effect of goal-type on the infliction of harm on the OG unlikely as an alternative explanation.

It is however possible, that participants were more likely to inflict harm on members of the OG in the minimal goal condition because of their intuitive understanding of the respective Public Good Game. Participants had no knowledge of the exact amount the initial

endowment or the payoff stemming from the Public Good when answering the dependent variables. This way any possible influence of the payoff participants missed out on because of the OG was excluded as an alternative explanation of the effect of goal-type on harm inflicted on the OG (Figure 4).

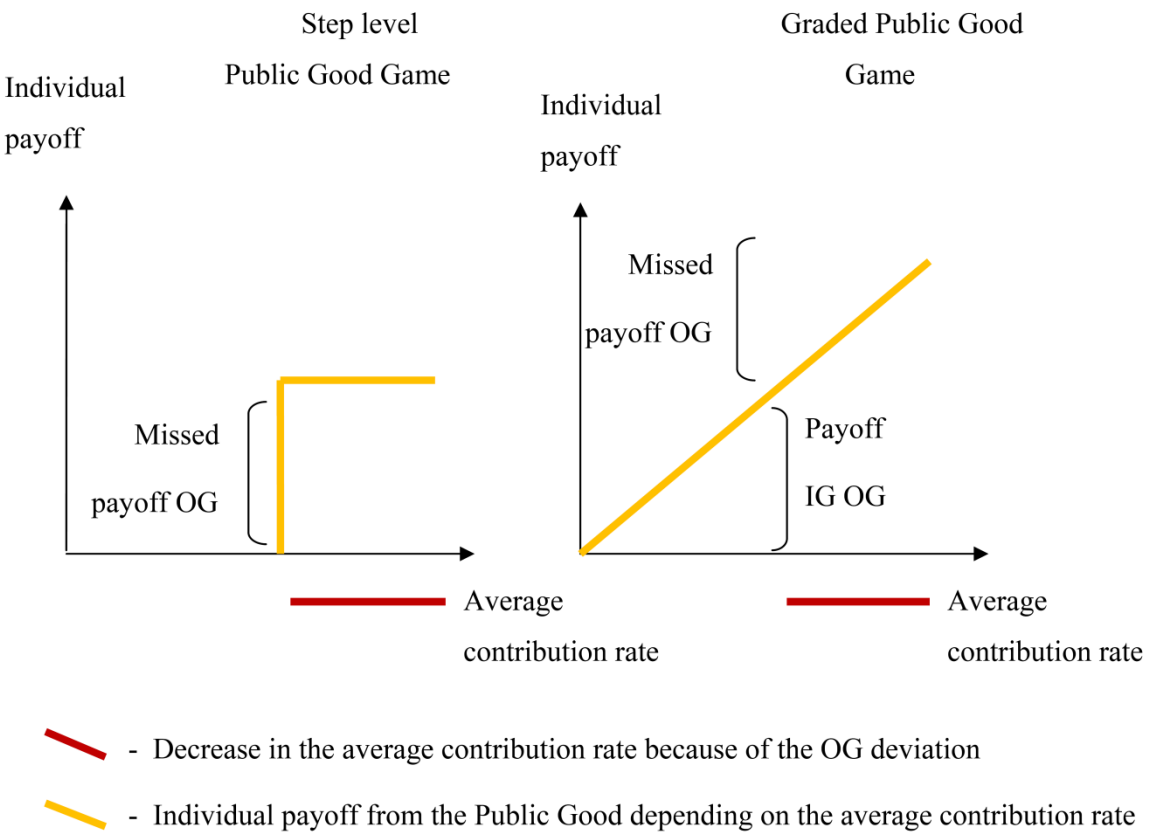

Figure 4: Conceptual Differences between Step-Level and Graded Public Goods Game

In the minimal goal condition there was no payoff from the Public Good (Figure 4: Payoff IG-OG) as a consequence of the OG deviation. This might have appeared as a more severe consequence in comparison to the maximal goal condition. Participants in the maximal goal condition knew they would receive some payoff from the Public Good as a result of the contributions of the IG players (Figure 4: Payoff IG-OG). An additional study is therefore needed that excludes structural differences in the payoff matrices as an alternative explanation for the effect of goal-type on the infliction of harm on members of a deviant OG.

4.4 Study IV: Goal-type Framing

4.4.1 Introduction

It is possible that participants in the minimal goal condition in Study 3 were more likely to inflict harm on members of the OG because of their intuitive understanding of the step-level public Goods Game. The possibility remains that the effect of goal-type on the infliction of harm was caused by a perceptual difference between minimal and maximal goal condition, even though there were no differences in payoffs between goal-type conditions in Study 3. Participants in the maximal goal condition received - according to the rules of a graded Public Goods Game - at least some payoff from the Public Good despite the deviation of the players of the OG. Participants in the minimal goal condition received no payoff from the Public Good because of the OG deviation (Figure 4). The fact that there was no payoff from the Public Good in the minimal goal condition might have appeared as a more severe consequence of the OG deviation in comparison to the maximal goal condition where participants received at least some payoff from the Public Good. Therefore, a change in the goal-type manipulation is required that excludes the possible influence of structural differences in the payoff matrices as an alternative explanation for the effect of goal-type on the infliction of harm on the OG.

4.4.2 Goal-type Manipulation

The specific aim of this study is to disentangle the manipulation of goal-type and the payoff matrices of graded and step-level Public Goods Game and. Therefore, participants in both conditions play a graded Public Goods Game. Participants´ representations of the superordinate goal as minimal or maximal are manipulated by a mere framing. The instruction in which participants are asked to suggest a contribution level of the initial endowment as superordinate goal is either framed as a maximal “the-more-the-better”-goal or a “black-and-white”-minimal goal. There are no functional implications of this framing. *Goal-Type Manipulation Check*. The assessment of participants´ goal-type representation showed that participants in the minimal goal condition scored significantly higher, $M = 3.71$; $SD = .99$, than participants in the maximal goal condition, $M = 2.65$; $SD = .95$, indicating a stronger minimal goal representation than participants in the maximal goal condition, $F(1, 46) = 13.94$, $p < .001$, $\eta^2 = .24$.

4.4.3 Group-goal Setting

As in the previous study no explicit goal was given in either condition. Instead the same group-goal setting mechanism as in Study 3 was used in both minimal and maximal goal condition. Participants could choose an average contribution level of 40 %, 60 % or 80 % as superordinate goal. This superordinate goal had - as in Study 3 - no functional implication. Participants engaged in both conditions in a graded Public Goods Game. *Contribution Decisions*. Participants in the minimal goal condition contributed on average 63 % of their initial endowment, *SD* = 19.84, participants in the maximal goal condition 75 %, *SD* = 14.10.

4.4.4 Methods

Design and Participants. Forty-seven students from the University of Jena participated in this study. The mean age of the participants was 21.36 years (*SD* = 2.56, range: 18-29), 28 of them female and 29 male. Twenty-four participants were randomly assigned to the minimal goal condition, 23 to the maximal goal condition.

Procedure. The basic experimental setup, the trial round, and the main round including the goal deviation of the OG were identical with the third study. The procedure differed only with regard to the number of average contribution levels that participants could choose from as superordinate goal. In this study only three contribution levels (40 %, 60 %, 80 %) were eligible. *Payoff.* Participants were not immediately informed about the exact amount of their payoff in order to exclude payoff feedback as a possible factor affecting participants´ replies to the dependent variables. Participants were informed at the very end of the study that their payoff from the second round was 2.10 €. The payoffs from both rounds summed up to 5.80 € .and were rounded to a final payoff for the participation in the study of 6 €.

4.4.5 Dependent Measures

Monetary Punishment. Social Exclusion. Ingroup Identification. Negative Emotions. The same measures of monetary punishment, social exclusion from future rounds, IG identification and negative emotions were used as in the previous studies. Both measures of harm, monetary punishment and social exclusion, were significantly correlated, r=.49, *p* = .001.

Ingroup Identification. Participants´ identification was significantly above the centre of the 7-point scale, $t(46) = 5.92$, $p < .001$; $M = 5.13$; $SD = 1.31$.

4.4.6 Results[23]

Goal-Type and Infliction of Harm. As in the first two studies participants in the minimal goal condition were more likely to punish members of the OG, $F(1,46) = 5.42$, $p = .012$, $\eta^2 = .11$, minimal goal: $M = 5.38$; $SD = 1.38$; maximal goal: $M = 4.35$; $SD = 1.64$, and exclude them from future rounds, $F(1, 46) = 5.04$, $p = .015$, $\eta^2 = .10$, minimal goal: $M = 4.79$; $SD = 1.79$; maximal goal: $M = 3.57$; $SD = 1.95$.

Goal-Type and Negative Emotions. Participants in the minimal goal condition were more likely to report higher levels of negative moral emotions toward the OG, $F(1,46) = 4.38$, $p = .021$, $\eta^2 = .09$, minimal goal: $M = 3.50$; $SD = 1.74$; maximal goal: $M = 2.52$; $SD = 1.44$.

A mediational analysis was carried out to examine the role of negative emotions as a mediator in the relationship between goal-type and monetary punishment of the OG. Goal type had a significant impact on negative emotions towards the OG, $\beta = -.30$, $p = .021$, and on monetary punishment of OG members, $\beta = -.33$, $p = .012$. Negative moral emotions were significantly related to monetary punishment of OG-members, $\beta = .54$, $p < .001$. When controlling for moral emotions, goal type had no longer a significant effect on the monetary punishment of OG-members, $\beta = -.18$, $p = .084$. In order to test the indirect effect bootstrapping (N=2000) was used to determine the standard error. The indirect effect of goal type monetary punishment of OG members by negative moral emotions was significant, BCa (95%) = [-1.1302, -.0730].

A second mediational analysis was carried out to examine the role of negative emotions as a mediator in the relationship between goal-type and the exclusion of OG members from future rounds. Goal type had a significant impact on the exclusion of OG members from future rounds, $\beta = -.32$, $p = .015$. Moral emotions predicted the exclusion of OG-members from future rounds, $\beta = .58$, $p < .001$. When controlling for moral emotions, goal type had no longer a significant effect on the exclusion of OG members from future rounds, $\beta = -.16$, $p = .11$. In order to test the indirect effect bootstrapping (N=2000) was used to determine the standard error. The indirect effect of goal type on the exclusion of OG members from future rounds via moral emotions was significant, BCa (95%) = [-1.3571, -.0270].

[23] All analyses with the independent variable goal-type are one-tailed analyses in this chapter.

4.4.7 Discussion

The main aim of this fourth study was a test of the goal-type hypothesis using a mere framing of the superordinate goal as minimal or maximal as a manipulation of goal-type. The independent variable goal-type - participants´ representation of the superordinate goal of contributing to a Public Good as either minimal or maximal - had been confounded with two specific payoff matrices in all previous studies. A graded Public Goods Game had been used to manipulate participants´ representation of the superordinate goal as maximal. A step-level Public Good Game had been used to manipulate participants´ representation of the superordinate goal as minimal. Study 3 had already controlled experimentally for payoff differences between minimal and maximal goal-type condition. However, a structural difference between step-level Public Goods Game (minimal goal condition) and graded Public Goods Game (Maximal goal condition) had remained in Study 3 as players did not receive any payoff from the Public Good in the minimal goal condition. It is possible that this lead to a perception of the OG deviation as more severe, resulting in a higher level of harm inflicted on the OG in the minimal goal condition.

Study 4 was designed to rule this structural difference between minimal and maximal condition out as an alternative explanation for effect of goal-type on the infliction of harm on members of a deviant OG: All participants engaged in Study 4 in a graded Public Goods Game - regardless of goal-type condition. Participants´ representation of the superordinate goal of making money by contributing to the Public Good as either minimal or maximal was manipulated with a mere framing of the group goal. An effect of goal-type on the infliction of harm of the deviant OG was found as in the previous studies. Participants in the minimal goal condition were more likely to inflict more harm on members of a deviant OG than participants in the maximal goal condition. This replication of the effect of goal-type on the infliction of harm allows the exclusion of a structural difference between payoff matrices of step-level Public Good Game and graded Public Good Game as an alternative explanation for the effect of goal-type on the infliction of harm on a deviant OG. The effect of goal-type on the infliction of harm on members of a deviant OG was - as in the previous two studies - mediated by negative emotions toward the OG.

4.4.8 Summary of Studies 2, 3 and 4

Studies 2, 3 and 4 showed that participants are more likely to inflict more severe harm on members of an OG deviating from a superordinate goal that is represented as a minimal goal

compared to the deviation from a superordinate goal that is represented as a maximal goal. Participants with a minimal representation of the superordinate goal showed also a higher level of negative emotions towards the OG. The effect of goal-type on punishment and exclusion of members of the OG was mediated by negative emotions toward the OG across Studies 2, 3, and 4. The incremental changes in the experimental design over these studies were supposed to rule out uncontrolled factors as explanations of the main effect of goal-type on the infliction of harm on the OG. Examples for these incremental changes in the experimental design are the selection of the level of the superordinate goal (Study 2 vs. Study 3) or the manipulation of goal-type by means of different payoff matrices (Study 3 vs. Study 4). A summarizing analysis across these three studies might show if the effects observed in all three studies add up to a consistent larger picture.

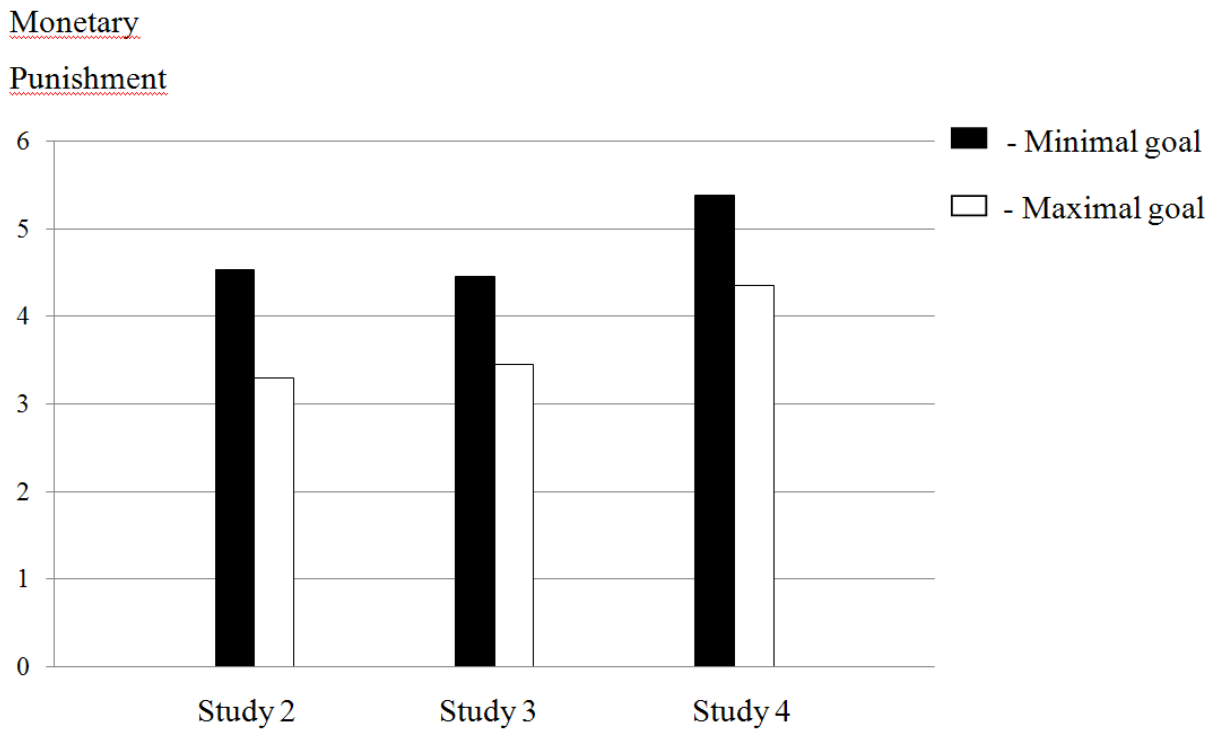

Figure 5: Goal-Type and Monetary Punishment: Studies 2 - 4

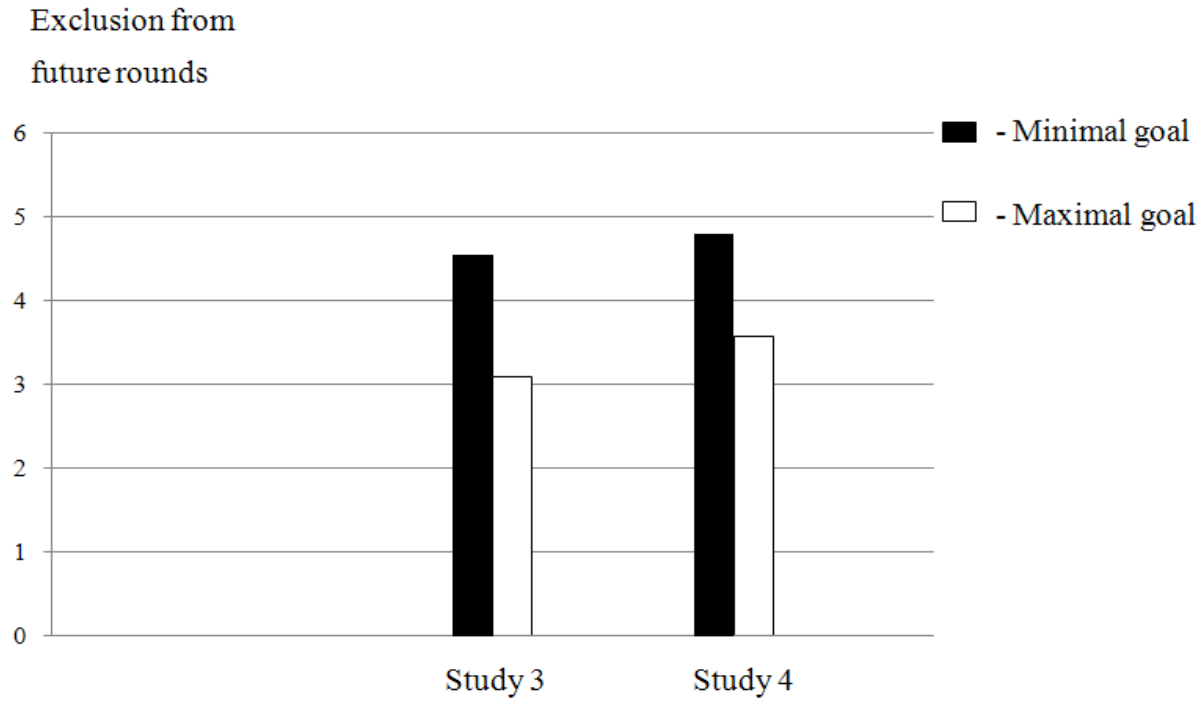

Figure 6: Goal-Type and Social Exclusion: Studies 3 and 4

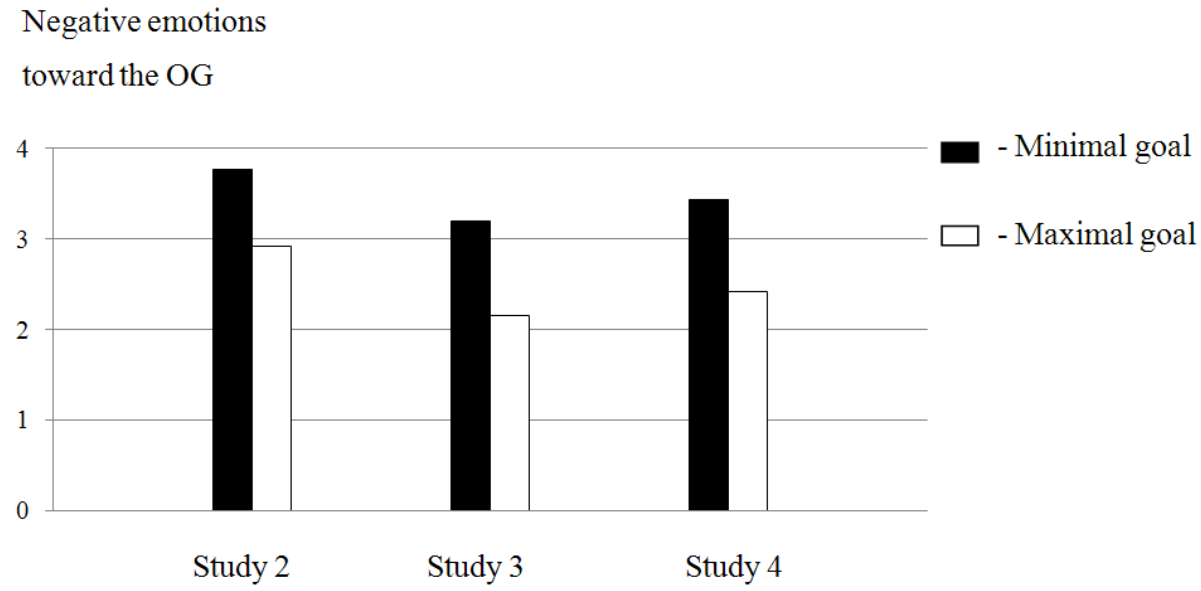

Figure 7: Goal-Type and Negative Emotions towards the OG: Studies 2 - 4

A MANOVA with the factors study and goal-type as independent variables and the dependent variables exclusion of the OG, punishment of the OG and negative emotions towards the OG was conducted. There were no interaction effects of study and goal-type[24], neither for the dependent variable monetary punishment, F (1,154) = .04, p < .959, nor for the dependent variable social exclusion from future rounds, F (1,154) = .94, p = .393.

A second MANOVA[25] with the independent variable goal-type showed across all three studies that participants with a minimal goal representation showed significantly more negative moral emotions towards the OG, F (1,154) = 13.24, p < .001, η^2 = .08, minimal goal: M = 3.56; SD = 1.78; maximal goal: M = 2.60; SD = 1.48, were more likely to exclude the OG from future rounds, F (1,153) = 9.58, p = .001, η^2 = .06, minimal goal: M = 4.50 SD = 2.08; maximal goal: M = 3.49; SD = 1.97, and punish them monetarily, F (1,154) = 14.31, p < .001, η^2 =.09, minimal goal: M = 4.76 ; SD = 1.88; maximal goal: M = 3.63; SD = 1.85. The two behavioral measures exclusion from future rounds a monetary punishment were significantly correlated, β = .575, p < .001.

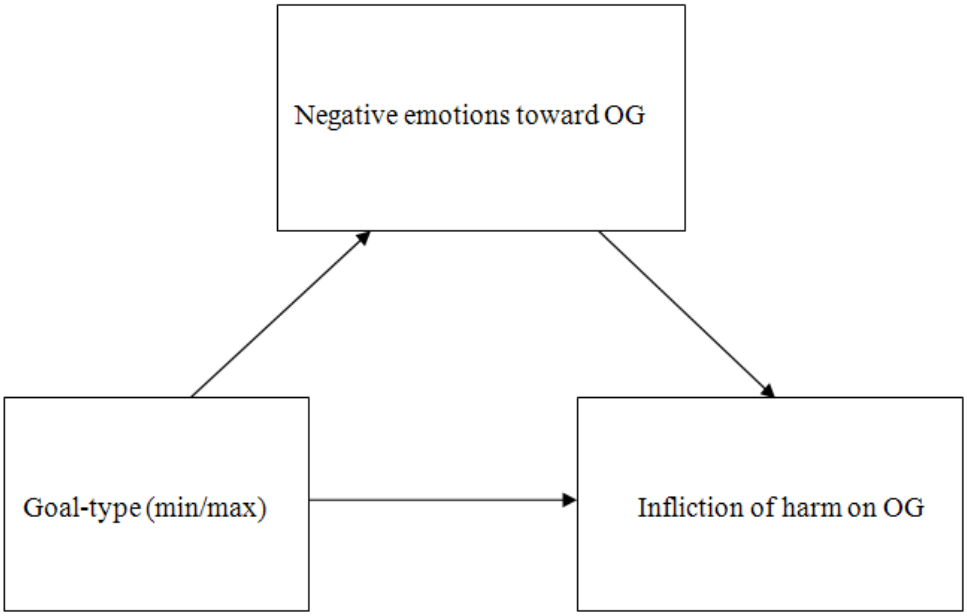

Figure 8: The Effect of Goal-type on the Infliction of Harm on the OG Mediated by Negative Moral Emotions towards the OG

[24] The factor study however did have a significant effect on the dependent variable monetary punishment, F (1,153) = 4.021, p = .010, η^2 = .05, one-tailed analysis,, due to a higher level of monetary punishment in both, minimal and maximal goal-type condition in Study 4.

[25] One-tailed analysis.

Two meditational analyses were carried out to examine the role of negative emotions as a mediator in the relationship between goal-type and the infliction of harm on the OG. In order to test the indirect effect bootstrapping (N=2000) was used to determine the standard error. The indirect effect of goal type on monetary punishment of OG members was significant, BCa (95%) = [-.9260,-.2388]. A second mediational analysis was carried out to examine the role of negative emotions as a mediator in the relationship between goal-type and exclusion from future rounds. The indirect effect of goal-type on the exclusion of OG members from future rounds was mediated by negative emotions towards the OG was BCa (95%) = [-1.0317,-.2970]. These analyses confirm the consistency of the mediation of the effect of goal-type on the infliction of harm by negative emotions toward the OG across the various studies.

Table 3: Table of Significant Effects of Goal-type on the Infliction of Harm Depending on Form of Goal-type Manipulation and Form of Harm

		DV: Form of harm inflicted on OG		
		Monetary punishment	*Social exclusion*	*Generalized social exclusion*
IV: Form of goal-type manipulation	*Payoff matrix*	Study 2 Study 3	Study 3	Study 3
	Framing	Study 4	Study 4	Study 5

This summarizing analysis shows a great deal of consistency across the last three studies. However, one minor drawback remains. Study 4 did not test an effect of goal-type on generalized social exclusion. Such a test of the effect of goal-type on generalized social exclusion would complete the permutation of goal-type manipulations and different forms of harm inflicted on the deviant OG (Table 3). This test of the effect of goal-type on the infliction of harm using at the same time the most subtle goal-type manipulation and the most

extensive measure of harm would conclude the test of both theoretical hypotheses and experimental paradigm. It is therefore the aim of a fifth study to show the effect of goal-type on the infliction of harm in form of generalized social exclusion using the framing technique presented in Study 4 to manipulate participants´ representations of the superordinate goal as either minimal or maximal.

4.5 Study V: Generalized Social Exclusion

4.5.1 Introduction

The aim of this fifth study is to replicate the effect of goal-type on the infliction of harm on a deviant OG that has already been found in the previous studies. The specific concern of this study is to manipulate goal-type with the framing technique that was already used in Study 4 and show the effect of goal-type for the specific case of generalized social exclusion. The subtle framing technique as goal-type manipulation in combination with the far-reaching infliction of harm on uninvolved members of the OG would complete the test of both experimental design and the goal-type hypothesis (Table 3). The same goal-setting technique was used as in Studies 3 and 4. The goal-type manipulation based on a framing of the group goal as either minimal or maximal and was identical to the goal-type manipulation used in Study 4.

4.5.2 Methods

Participants. Forty-seven students from the University of Jena participated in this study. Twenty-eight participants were female, 19 male. The mean age was 21.53 years, *SD* = 2.48, range: 18-31**.**

Procedure. The procedure of this study was identical to the procedure used in study 4.
Payoff main round. Participants´ payoffs were kept constant across conditions and goal-levels as in Study 4. All participants received 3.80 € as payoff from the first round and 2,10 € from the second round. The total payoff of 5.90 € was rounded to 6 €.

4.5.3 Measures

Generalized Social Exclusion. The general concept of the item of generalized social exclusion was similar to the measure used in Study 3. It was assessed after the actual computer-based study as a separate study in a pen-and-paper format. Participants were told a cover-story that explained why they had to select a number of participants from previous studies that were not invited to participate in a future study. One improvement of this measure of generalized social exclusion over the one used in Study 3 was that this measure was formulated as an exclusion of individuals from future studies, and not as a non-inclusion[26]. It

is important to point out that there was no given number participants had to exclude. In addition there was no maximum number of individuals that could be excluded, resulting in inter-individual variance in the number of excluded individuals.

Ingroup Identification. Participants´ identification with their IG was measured using the same items as in the previous studies. Participants´ identification with their IG, $t_{1:}$ $M = 4.98$; $SD = 1.51$ and t_2: $M = 5.25$; $SD = 1.43$, was both in t_1 and t_2 significantly higher than the centre of the scale, $t_{1:}$ $t(46) = 4.45$, $p < .001$ and $t_{2:}$ $t(46) = 6.00$, $p < .001$. There was neither in t_1, $\beta = -.022$, $p = .882$, nor in t_2, $\beta = -.060$, $p = .687$, a main effect of goal-type on the level of identification with the IG. *Goal-type Manipulation Check.* Three items assessed on a seven-point scale participants´ representation of the goal-type, e.g. "The average contribution level of […] must be achieved"; "The average contribution level of […] should ideally be achieved"; Cronbachs $\alpha = .67$. The items were adjusted to the contribution level participants had chosen previously. Participants in the minimal goal condition scored significantly higher on the goal-type manipulation check, $F(1, 46) = 15.06$, $p < .001$, $\eta^2 = .25$, indicating a stronger minimal goal representation than participants in the maximal goal condition, minimal goal: $M = 4.91$; $SD = 1.33$; maximal goal: $M = 3.33$; $SD = 1.45$. *Contribution Decisions*. Participants in the minimal goal condition contributed on average 79 %, $SD = 16.5$. Participants in the maximal goal condition contributed on average 68 %, $SD = 18.6$. The set group goal for the average contribution level and participants actual contribution decisions were significantly correlated, both in t_1, $r = .72$, $p < .001$, and t_2, $r = .58$, $p < .001$.

[26] The cover story was that participants of former studies that had applied to participate in a future well paid study had to be selected from a database. Not all participants from that database could participate. Participants of this study therefore had to select participants from the database that would not be invited to participate in the future study. Participants in the present study as uninvolved participants were asked to select a number of players that could not be invited to the future study in order to achieve an unbiased selection process. Former participants were listed symbolized by a code of letters and numbers. Although participants of the alleged earlier studies were kept anonymous on the list, their group-identity as either members of the beta- or gamma-group was evident. The code of 10 participants from the database contained the letters "be" - indicating a beta-group membership - and the code of 10 participants contained the letters "ga" indicating a gamma-group membership.

4.5.4 Results

Goal-type and Generalized Social Exclusion. Participants in the minimal goal condition were not more likely to exclude uninvolved participants in general - regardless of their group membership -from the participation in future profitable studies, $F(1, 46) = 0.01$, $p = .908$, minimal goal condition: $M = 7.12$; $SD = 2.88$; maximal goal condition: $M = 7.23$; $SD = 3.44$. Participants in the minimal goal condition were not significantly more likely to exclude members of the OG from the participation in future profitable studies, $F(1, 46) = 1.37$, $p = .248$, minimal goal condition: $M = 4.16$; $SD = 2.21$; maximal goal condition: $M = 3.41$; $SD = 2.18$.

If however the number of excluded OG members is set into relation to the absolute number of excluded participants - thus controlling for the inter-individual variance in the number of excluded individuals - a difference between minimal and maximal goal condition becomes evident. Participants in the minimal goal condition were more likely to exclude more OG members than IG members from the participation in future profitable, $F(1, 46) = 4.58$, $p = .019$, $\eta^2 = .09$[27], share of excluded OG members in the minimal goal condition: $M = 62$ %; $SD = 26$; share of excluded OG members in the maximal goal condition: $M = 47$ %; $SD = .22$. The relation between excluded OG members and excluded IG members deviated only in the minimal goal condition from an exclusion by chance of 50 % excluded OG-members and 50 % OG-members, minimal goal condition $t(24) = 2.23$, $p = .018$, maximal goal condition $t(21) = -.74$, $p = .233$.

4.5.5 Discussion

It was the aim of the study not only to replicate the effect of goal-type on the infliction of harm on a deviant OG. But to do so using a mere framing as goal-type manipulation and to show the effect of goal-type on generalized social exclusion as a specific form of harm inflicted on members of the OG. . One improvement of the measure of generalized social exclusion in this study - in contrast to Study 3 - was that it was explicitly phrased as an exclusion item.

There was no effect of goal-type on the general tendency to exclude participants regardless of their group membership. Similarly participants in the minimal goal condition were not more likely to exclude OG members in general from future profitable studies. This lack of difference between participants in minimal and maximal goal condition however can

[27] One-tailed analysis

be explained by a methodological feature of the measure for generalized social exclusion in this study. The measure of generalized social exclusion used in this study was not based on a quantitatively specified scale. Participants could choose how many participants from former studies they wanted to exclude from the participation in highly profitable future studies. A second analysis controlled for any inter-individual differences in terms of the absolute number of participants excluded from the future study by analyzing the quotient of excluded OG members and the total number of excluded participants. This analysis reveals a different picture. Participants in the maximal goal condition made no difference between excluding IG and OG members. They excluded members of both groups with the same probability. Participants in the minimal goal condition however deviated in their exclusion decisions significantly from chance. Participants in the minimal goal condition were significantly more likely to exclude members of the OG than members of the IG. This result is clearly in line with the results of the previous studies.

5 General Discussion

5.1 Summary

The present research set out to pursue two central research goals. The first goal was a test of the *goal-type hypothesis*. It was hypothesized that the deviation from a minimal superordinate goal would lead, because of its dichotomous structure, to a higher level of harm inflicted on members of a deviant OG than the deviation from a maximal, graded superordinate goal (*goal-type hypothesis*). A second additional hypothesis had been derived. It was hypothesized that this effect of the goal-type of a superordinate goal on the infliction of harm on members of a deviant OG would be mediated by negative moral emotions towards the deviant OG (*negative emotions hypothesis*). All studies were set in a specifically designed experimental paradigm based on Public Goods Games. It was the second main goal of the present research to develop this experimental paradigm in order to test these hypotheses. There were two main requirements for this experimental paradigm. First, it was supposed to allow the measurement of harm on members of a deviant OG in the form of actual behavior. The second main requirement for the research paradigm was a strictly experimental design in order to exclude external factors as alternative explanations for an effect of goal-type on the infliction of harm on members of a deviant OG.

Five studies were conducted to test the two main hypotheses. Both hypotheses could be confirmed. Participants with a minimal goal-type representation of a superordinate goal were more likely to inflict more harm on members of an OG deviating from this superordinate goal than participants with a maximal goal-type representation. This effect was independent of the way goal-type was manipulated and was found for three different measures of harm including a measure of generalized social exclusion. This effect of goal-type on the infliction of harm on members of a deviant OG was mediated by negative moral emotions toward the OG. This mediation was found independent of the way goal-type was manipulated and for all measures of harm. The following chapters will discuss the evolution of the experimental paradigm across studies, set the results of these studies in a broader context and discuss future research directions.

5.1.1 Goal-setting Procedure

The superordinate goal of making money by contributing collectively to a Public Goods Game was identical across all studies. However, two different procedures were used to introduce this superordinate goal. In the first two studies players received no information about their co-players except for their membership in one of the arbitrary group categories. Participants had no information if the goal of contributing to the Public Goods Game was shared by their co-players previous to their own contribution decision. However, trust in other-players behavior is crucial for cooperation in a Public Goods Game to occur. Therefore a test-round was introduced in all Studies from Study 3 on to communicate that the superordinate goal of contributing to the Public Goods Game was a shared goal among IG members. This first round established in addition not only that the superordinate goal was known to all IG members, but that the other IG members would also behave in line with this goal and contribute a share of their initial endowment to the Public Good. This introduction of the superordinate was identical for all players regardless of goal-type condition (minimal or maximal superordinate goal).

5.1.2 Goal-type Manipulation

Two different manipulations of participants´ representation of the superordinate goal as either minimal or maximal were used. A first goal-type manipulation was based on slightly different designs of a Public Good Game (Studies 1 - 3). A Graded Public Goods Game was used to resemble the graded structure of a maximal superordinate goal. The superordinate goal consisted in this maximal goal condition in an ideal average contribution rate of 100 % the initial endowment resulting in a maximum payoff on a collective level. A step-level Public Goods Game was used to resemble the dichotomous “black/white” structure of a minimal goal. A threshold of an average contribution level had to be reached in order to receive a payoff from the Public Good. However, this manipulation of goal-type was potentially flawed. Participants received - as a consequence of the OG´s deviation - a slightly lower payoff from the step-level Public Goods Game (minimal goal condition) than from the graded Public Goods Game (maximal goal condition). Study 3 controlled experimentally for this difference in payoffs. However, the structural differences between step-level Public Goods Game and graded Public Goods Game remained. In the maximal goal condition players received at least some payoff from the Public Good. In the minimal goal condition participants received no payment at all from the Public Good. This might have lead to a perception of the OG´s deviation as more severe, even though there were no differences in

final payoffs. Therefore, a second manipulation of goal-type was based on a mere framing (Studies 4 - 7). All players engaged in a graded Public Goods Game. The superordinate goal was either framed as a graded "shades-of-grey"-goal or a minimal "black-or-white"-goal.

5.1.3 Goal-type and the Infliction of Harm

The infliction of harm was measured on a variety of measures, ranging from the deduction of money from the OG, to the exclusion of members of the OG from future rounds of the Public Goods Game and the generalized social exclusion of uninvolved members of the OG from the participation in future profitable studies. The *goal-type hypothesis* was tested using both kinds of goal-type manipulation for all dependent measures of harm. Participants were more likely to inflict more harm on members of an OG deviating from a minimal superordinate goal than members of an OG deviating from a maximal superordinate goal in all studies. The effect of goal-type on the infliction of harm on a deviant OG was replicated various times across both goal-type manipulations and for various measures of harm. The effect of goal-type on the infliction of harm on a deviant OG was independent of the kind of goal-type manipulation.

5.1.4 Outgroup Goal-deviation

It was the aim of the present research to scrutinize the impact of a representation of a superordinate goal as either minimal or maximal on the infliction of harm on members of a deviant OG. Therefore a clear manipulation of the OG´s deviation was designed in the framework of the experimental design. The OG ´s deviation from the superordinate goal of making money by contributing to a Public Goods Game consisted in a contributing significantly below the level of a superordinate goal and less than the members of the participants´ IG. All members of the OG contributed less than the members of the IG. This pattern of the OG´s deviation from the superordinate goal was chosen deliberately to create an unambiguous case of OG deviation. The contribution decisions of both IG members and OG members were staged in order to keep the OG-deviation pattern identical for all participants. However, this clear-cut manipulation of the OG´s deviation from the superordinate goal has a theoretical drawback as only deviant OG-members of the OG were punished respectively excluded. But the great destructive potential of the infliction of harm in an intergroup context originates in particular from the generalized infliction of harm on uninvolved members of the OG. Future research should therefore address the infliction of harm on uninvolved group members by disentangling the deviation of OG members and the infliction of harm on members of this OG. One such OG-deviation pattern in the framework of the present experimental paradigm could be the partial deviation of the OG (2 of a total 3 OG members deviating significantly from the superordinate goal, 1 OG member behaving in line with the superordinate goal). The aim of this heterogeneous OG deviation pattern would be to show the infliction of harm even on members of the OG that are not deviating from the OG. However, the present research started already to address this concern by assessing the effect of goal-type on the generalized exclusion of uninvolved OG members.

5.2 Generalization

Study 3 measured participants´ inclination to exclude uninvolved members of IG and OG from the participation in future profitable studies. This effect of goal-type on generalized social exclusion is different from the two other measures of harm. Both monetary punishment and exclusion from future rounds were directed at the very individuals deviating from the superordinate goal of contributing at a certain level to a Public Goods Game. The members that were excluded from the participation in future profitable studies via generalized social exclusion however had not deviated from this superordinate goal. Their only connection to the deviant members of the OG was their abstract, arbitrary OG membership. It is important to

point out that participants had in none of these 3 studies an opportunity to profit from their exclusion decision as they themselves had no possibility to participate in these future profitable studies. Participants were more likely to include IG members instead of OG members to participate in these future studies. Study 5 similarly assessed participants´ tendency to exclude uninvolved IG and OG members from future studies using a mere framing manipulation of goal-type. Participants in minimal and maximal goal condition did not differ in the absolute number of excluded IG and OG members. But only participants in the minimal goal condition deviated from a fairness strategy when excluding uninvolved IG and OG members. Participants in the maximal goal condition were equally likely to exclude IG and OG members. Only participants in the minimal goal condition were significantly more likely to exclude uninvolved OG members from the participation in future studies. But why were participants with a minimal goal representation of a superordinate goal more likely than participants with a maximal goal representation to inflict harm on uninvolved members of the deviant OG? The following paragraphs present 3 distinct attempts to explain this particular effect of goal.-type on the generalized exclusion of OG members.

5.2.1 Generalization and Experimental Paradigm

The effect of goal-type on the generalized exclusion of harm could be based on the same process as the effect of goal-type on monetary punishment and the exclusion of OG member from future rounds of the Public Goods Game. The measure of generalized social exclusion was significantly correlated with these two more direct measures of harm in Study 2. The effect of goal-type on generalized social exclusion was in addition mediated by negative moral emotions toward the OG, just as the effect of goal-type on the two other measures of harm. However, the measure of generalized social exclusion differs decisively from the two other measures of harm in terms of its effect of members of the OG that were merely connected to the small subgroup of OG deviants by an arbitrary and trivial group membership. The argument that the effect of goal-type on generalized social exclusion was based on the same process as the effect of goal-type on monetary punishment and social exclusion from future rounds would imply that a strong generalization process from OG sample (the 3 deviant OG members in the Public Goods Game) to OG population (all OG members) was present in both minimal and maximal goal condition.

Research on the generalization of sample information in intergroup contexts has reported that sample information that is favourable from an IG perspective is directly generalized to the target population reflecting the sample distribution (Doosje, Spears, & Koomen, 1995).

However, sample information that is unfavorable from an IG perspective is not always generalized directly to the target population. Instead identity protection strategies are used such as a disproportionate generalization from small or heterogeneous samples. If the sample does not allow such a biased generalization other strategies are used to protect a positive social identity. In the case of a homogeneous or large sample the - from an IG perspective - unfavourable sample information might be generalized to the target population. However, in this case individuals might perceive the population as more variable in order to protect their social identity. In the present research the sample information was represented by participants´ contribution decisions in the Public Goods Game. The valence of this sample information was given by the superordinate goal of making money by reaching a certain average contribution level. The samples were small as only 3 IG members and 3 OG members engaged in the Public Goods Game. The sample information was homogeneous. The OG contributed in all studies significantly below the average contribution level of the superordinate goal and at the same time significantly below the contribution level of the IG. The sample information of the IG and OG was also distinct as contribution patterns of IG and OG never overlapped (the lowest contributing IG member always contributed more than the highest contributing OG member). The sample information was therefore favorable from an IG perspective not necessitating one of the described identity protection strategies. It is therefore highly likely that the sample distribution was generalized to the population level: All OG members deviated significantly from the superordinate goal. At the same time all IG members were contributing in line with this superordinate goal. Thus, it can be expected that OG members at a population level - all existing OG members regardless if they participated in the study - were considered to be equally deviant as the sample group of OG members.

This generalization tendency might have been further amplified by the experimental context. The intergroup context of the present research is based on the Minimal Group Paradigm (Tajfel et al., 1971). Participants had no possibility for face-to-face interaction or communication. All participants remained anonymous. The group categories were arbitrary and abstract. There was no interaction history with members of the OG prior to the experiment. The OG´s deviation from the superordinate goal was the only sample of information that could be attributed to a membership in the OG. This experimental control of sources of individuating information is highly likely to increase the homogeneity perceptions of an OG (Ostrom & Sedikides, 1992). Both the quality of the sample information as well as the specific design of the experimental paradigm favored therefore the generalization of the sample information - the OG´s deviation from the superordinate goal - to the population level

of all OG members. Thus suggesting that all OG members are deviants from the superordinate goal, justifying their punishment regardless if they had actually committed such a deviation or not.

This first attempt to explain the effect of goal-type on the generalized social exclusion of uninvolved OG members assumes in other words a very high level of generalization from OG information sample to OG population level in both minimal and maximal goal condition. The effect of goal-type on all 3 measures of harm would - according to this rationale - be based on the same processes. However, this approach cannot explain the result of Study 5 that only participants in the minimal goal condition deviated in their infliction of generalized social exclusion on IG and OG members from a fairness strategy. Two additional explanation attempts suggest in contrast a distinct effect of goal-type on the generalization process compared to the two direct, "retributive" measures of harm monetary punishment and exclusion from future rounds. Both approaches are based on the observation that the generalization process is likely to be closely related to perceptions of OG homogeneity. The higher the perceived homogeneity of the OG with regard to a trait - such as the deviance from the superordinate goal - the stronger the generalization of OG sample information to an OG population level. Both alternative explanation attempts for the effect of goal-type on the generalized exclusion of uninvolved OG members assume therefore an effect of goal-type on the perception of OG homogeneity. Both approaches suggest that a minimal goal representation leads to a higher level of perceived OG homogeneity than a maximal goal representation.

5.2.2 Goal-type, Threat and Generalization

The deviation of an OG from a minimal superordinate goal is expected to be perceived as more severe and negative than the same deviation from a maximal superordinate goal. Research by Henderson-King and Nisbett demonstrated that the generalization from individual OG-members to the population level of the OG is particularly likely to occur for negative, stereotypical information (Henderson-King & Nisbett, 1996). Similarly research in the framework of the Group Attribution Error paradigm could show that participants inferred more homogeneous attitudes among members of threatening compared to non-threatening groups (Corneille, Yzerbyt, Rogier, Buidin, 2001). This line of reasoning would suggest that an OG´s deviation from a minimal superordinate goal is more likely to be perceived as negative and threatening compared to an OG´s deviation from a maximal superordinate goal. A more negative evaluation and a higher level of perceived threat should in turn lead to a

higher level of perceived OG homogeneity and a stronger generalization from an OG sample distribution to the level of the OG population.

5.2.3 Goal-type as a Mindset and Generalization

An alternative explanation for the effect of goal-type on generalized social exclusion that is also based on an effect of goal-type on perceived OG homogeneity follows a rather cognitive rationale. Previous research has conceived the concept of goal-type beyond the representation of a superordinate goal as a mindset (Berthold et al., 2011). Mindsets are considered as cognitive procedures related to how one chooses between various alternatives (see Gollwitzer, Heckhausen, & Steller, 1990; Galinsky & Moskowitz, 2000). This concept of goal-type as a mindset suggests a further reaching effect of goal-type with the potential to affect the perception of the OG´s homogeneity in particular (e.g. Ostrom & Sedikides, 1992). A minimal goal mindset with its dichotomous structure may lead to a higher level of perceived OG homogeneity than a graded maximal goal mindset, which allows a more differentiated perspective on the OG. This higher level of perceived OG homogeneity should in turn lead to a more pronounced generalization of sample information regarding the OG such as the deviation from a superordinate goal of some OG members to an OG population level. This line of reasoning suggests that the effect of goal-type on the generalized social exclusion of uninvolved OG-members is mediated by different levels of perceived OG homogeneity, independently of the perceived level of threat. Participants with a minimal "black-and-white"- mindset are expected to perceive the OG as more homogeneous than participants with a maximal "shades-of-grey"- mindset. A perception of the OG as more homogeneous should facilitate the generalized infliction of harm on uninvolved OG-members. Again, the level of perceived OG homogeneity is expected to be strongly related to the extent of generalized social exclusion of uninvolved OG members.

Further research is required to substantiate the presumption that the effect of goal-type is based on a specific effect of goal-type on the generalization of OG sample information to an OG population level. Future studies should therefore test the assumed effect of goal-type on the perception of OG homogeneity. The OG sample information in the present research was very homogeneous (very similar OG contribution decisions, high distinctiveness from IG contribution decisions, suppression of individuating information). More heterogeneous OG sample information should avoid a constantly high level of generalization across goal-type conditions and allow a better test of a possible effect of goal-type on the perceived OG homogeneity. In addition OG homogeneity could be measured twice, once right after the goal-

type manipulation (t_1) and once after the deviation of the OG from the superordinate goal (t_2) in order to determine if a possible effect of goal-type on the perceived homogeneity of the OG is based on different levels of OG threat in minimal and maximal goal condition. An effect of goal-type on perceived OG homogeneity in t_1 would support the rationale based on different mindsets in minimal and maximal goal condition. An effect of goal-type on perceived OG homogeneity in t_2 but not in t_1 would favor in contrast an explanation based on different levels of perceived threat depending on goal-type of the superordinate goal.

5.3 Negative Moral Emotions

The "negative emotions" hypothesis suggested that participants report different levels of negative emotions in response to the OG´s deviation from a superordinate goal depending on the representation of this superordinate goal as either minimal or maximal. This hypothesis could be confirmed. Participants in the minimal goal condition were more likely to report more negative moral emotions toward the OG than participants in the maximal goal condition in Studies 2, 3, and 4. Negative moral emotions mediated the effect of goal-type on the infliction of harm on members of a deviant OG in Studies 2, 3, and 4. The "negative emotions" hypothesis addresses the effect of goal-type on negative moral emotions toward the OG on a quantitative dimension. Negative emotions were measured in Studies 2, 3 and 4 by three items, each addressing one specific emotional reaction toward the OG (anger, contempt and disgust). These three items were significantly correlated in all studies. They were therefore consolidated into one combined construct of negative moral emotions.

But various approaches to moral emotions have highlighted the importance of a qualitative differentiation between various distinct types of moral domains. Rozin and co-authors suggested in their *CAD Triad Hypothesis* that three basic negative other-regarding emotions - contempt, anger and disgust - are elicited by violations of three basic moral codes - community, autonomy and disgust (Rozin, Lowery, Imada, & Haidt, 1999). This *CAD-Triad Hypothesis* links contempt to the moral code of community, anger to the moral code of autonomy and disgust to the moral code of divinity. The CAD Triad-Hypothesis suggests a qualitative fit between different moral domains and matching emotional responses of deviations in this respective moral domain. The superordinate goal of collectively contributing to a Public Good may best be categorized as belonging to the *moral code of autonomy* as it was based on a contribution level that players selected themselves. This superordinate goal of contributing to a Public Good lacked in addition any features that would have been typical for the *moral codes of community* or *divinity* such as hierarchical structure

or a relation to the abstract concept of purity. Participants´ emotional response to the OG´s deviation was measured in studies 2, 3, and 4 - as described earlier - on three items, each targeting anger, contempt or disgust. These three items were significantly correlated with each other, forming a homogenous scale. But the three emotional items differed consistently with regard to the level of their means. Participants were inclined to report much higher levels of anger than contempt or even disgust. This pattern might therefore be interpreted as a fit between the moral code of the superordinate goal of contributing in a Public Goods Game and anger as the most appropriate emotional reaction to a deviation from this superordinate goal. However, this assumption is at the given point purely speculative. But it opens up a new interesting perspective on the role of negative moral emotions in the relation between the goal-type of a superordinate goal and the infliction harm on deviants from this goal. It raises the questions if the effect of goal-type on the infliction of harm on deviants is mediated by specific moral emotions depending on the specific moral code of the superordinate goal.

Examples:

The present research suggests that the effect of goal-type on the extent of harm inflicted on members of a deviant OG is mediated primarily by anger if the deviation from the superordinate goal is mainly perceived as a frustration or blockage of the group´s goal as it may be the case in the present research. But it might be presumed that the effect of goal-type on the infliction of harm is rather mediated by contempt if the superordinate goal is rooted in the moral code of community, that is characterized by a*uthority* and *respect* (e.g. in a highly stratified and hierarchical context such as the military. The deviation from an order by a group with lower status may result in various degrees of punishment depending on the representation of the order as minimal or maximal. The primary emotional reaction mediating this effect ought to be contempt). The effect of goal-type on the infliction of harm on a deviant OG in a context rooted in the moral code of divinity and purity should result in yet another primary emotional reaction, moral disgust (e.g. the Catholic Church has a strict policy against abortion. A group of Catholics may support the practice of abortion in their commune under certain circumstances, for example if the pregnancy is a result of rape. Other Catholics may object this practice and demand punishment and possibly the exclusion of this deviant OG from the SOG, the Catholic Church. The severity of this punishment is expected to depend on their representation of the superordinate goal of protecting all life, born or unborn, as either minimal or maximal. It may be assumed that the predominant emotion mediating this

effect of goal-type on the infliction of harm is moral disgust, because the superordinate goal is set in the moral code of *divinity*).

The *CAD Triad Hypothesis* suggests in sum that the effect of goal-type on the infliction of harm is mediated by *contextually specific* moral emotions. However, a different approach argues that specific negative emotions are related to particular *social functions*. Fischer and Roseman proposed that for example anger is prototypically related to short-term attacks but aims from a long-term perspective at potential reconciliation (Fischer & Manstead, 2008; Fischer & Roseman, 2007). The social function of anger is to alter another person´s undesirable behavior with the use of aggression, but to do so in order to improve the entire relation in a long-term perspective. Contempt on the other hand is related to short-term derogation, but also to long-term social distancing. It has been argued similarly that moral disgust - disgust elicited by social transgressions such as lying, cheating and stealing opposed to forms of pathogen disgust or sexual disgust - is also related to the social exclusion of deviants (Tybur, Lieberman, & Griskevicius, 2009). Moral disgust is - from a social functional perspective - targeted at the prevention of norm deviating behavior that threatens to inflict direct cost via defection or an indirect cost by disrupting cooperation in a group and the group´s cohesion (Cotrell & Neuberg, 2005).

It may be suggested accordingly that minimal and maximal superordinate goals serve different social functions at a group level, regardless of the moral code of this superordinate goal. *Maximal superordinate goals* are characterized by graded evaluations in relation to an ideal that group members should strive for. It is therefore suggested that the social function of a maximal group goal is to define the prototype of a social category. *Minimal superordinate goals* on the contrary are defined by a dichotomous evaluation structure. A minimal goal is either fully achieved or not at all. It is therefore suggested that the social function of a minimal group goal is to define the borders of a social category. The deviation from a maximal superordinate goal activates a social relational goal of incentivizing the deviants to adhere more closely to this maximal goal. The behavioral reactions toward the deviants ought to reflect this social relational goal (example: If the superordinate goal of striking is represented as a maximal goal, deviants - workers who do not strike - may be faced with relatively mild forms of punishment in reaction to their deviance such as a fee by their union or name calling and appeals by their colleagues to join the strike). The primary emotional reaction in response to a deviation from a maximal goal should be anger. The deviation from a minimal superordinate goal activates in contrast a distinct social relational function that is

characterized by social distancing. The corresponding form of harm toward the deviants would be the cognitive and behavioral exclusion from the superordinate group (example: If the superordinate goal of striking is represented as a minimal goal, deviants may be no longer considered as part of the superordinate group "workers" by their striking co-workers. They may be faced with ostracism - co-workers refusing to work with them - and even physical violence and exclusion). Participants should not only report anger in response to a deviation from a minimal superordinate goal, but also contempt and moral disgust.

The effect of the goal-type of a superordinate goal on the infliction of harm could be interpreted in support of these assumptions about the social functions of minimal and maximal group goals. Participants with a minimal goal representation of a superordinate goal were not only more inclined than participants with a maximal goal representation to punish deviants, to exclude them from future interactions or to exclude uninvolved members of the OG who were only connected to the deviants by an arbitrary group membership. However, despite the effect of goal-type on the quantity of negative moral emotions towards the OG, no effect of goal-type on the quality of these negative moral emotions could be found in Studies 2 - 4. Participants in the minimal goal condition were not systematically more likely to report higher levels of contempt and disgust than participant in the maximal goal condition. This finding might be attributed to the fact that the measurement of negative moral emotions in the present research was not targeted at the detection of qualitative but quantitative differences between conditions (e.g. participants were not asked to identify the most fitting emotional reaction; the qualitative differences between contempt, anger and disgust were in addition not made salient, compare Rozin et al., 1999). Future research should therefore not only measure the social relational goals that are activated by the deviation by an OG from a minimal respectively maximal superordinate goal, but also focus on the quality of the emotions that are triggered by this deviation.

Finally it has to be pointed out that *CAD Triad Hypothesis* and the *social functional approach* to moral emotions are far from incompatible as both approaches apply to different levels of analysis. The *CAD Triad Hypothesis* is concerned with the situational appraisal of specific emotions. The *social functional approach* is in contrast based on relational goals - both on an interpersonal as well as intergroup level - and especially the activation of responding behavioral reactions. But negative moral emotions are by no means exclusive. It is actually highly likely that individuals experience more than one negative moral emotion in response to a perceived transgression. A symbiosis of both approaches for the case of minimal

and maximal group goals could imply for example that the moral domain of a transgression sets a "background" emotion - contempt, anger, disgust - as a default for any reaction to a perceived deviation from a superordinate goal. The goal-type of this superordinate goal may determine in turn the social relational goal for the reaction toward an OG deviating from this superordinate goal: Short-term aggression and anger in case of a maximal goal representation with the option of long-term reconciliation in the case of a successful "conversion" of the deviants towards a closer adherence to the superordinate goal; and social distancing, contempt, disgust and social exclusion of the deviants in case of a minimal goal representation of the superordinate goal.

These assumptions are however highly speculative. Yet, negative moral emotions play a crucial role in the relation between goal-type of a superordinate goal and the infliction of harm on deviants from this goal as the present research has shown. A further analysis should focus therefore on the activation of specific negative moral emotions and their related social functions and consequentially on their capacity to activate qualitatively different forms of harm inflicted on deviants.

5.4 The Impact of Goal-type on the Influence of Attenuating Factors

The importance of negative moral emotions in the relation between the goal-type of a superordinate goal and the extent of harm that is inflicted on members of a deviant OG has been shown. The different structures of minimal and maximal goals lead to different evaluations of the OGs deviation and elicit different levels of negative moral emotions toward the deviants. The extent of negative emotions strongly influences the extent of harm that is inflicted on members of the deviant OG.

However, it may be suggested that this is not the only implication of the representation of a superordinate goal as either minimal or maximal for the infliction of harm on a group of deviants. It has been argued that the extent of harm that is inflicted in response to an offense is not only a function of the perceived magnitude of harm caused by this offense but also the perception of mitigating or exacerbating circumstances (Carlsmith, Darley, & Robinson, 2002). But the goal-type of a superordinate goal has also the capability to systematically affect this perception of attenuating respectively extenuating circumstances. One important factor that is moderating the infliction of harm in response to an offense is the attribution of the transgression as a result of either internal or external causes (Weiner, 1986). It is argued accordingly that the representation of a superordinate goal as a maximal goal allows an

external attribution of an OG´s deviation from this superordinate goal while a minimal goal-representation facilitates an internal attribution pattern. The graded structure of a maximal goal permits an external attribution of an OG´s deviation (example: "Did the OG deviate from the average contribution goal of 80 % because they did not understand the game? They didn´t contribute at goal level, but they contributed at least some of their initial endowment..."). The relative evaluation that coincides with a maximal goal attribution allows giving the members of the deviating OG the "benefit of the doubt", thus attenuating the extent of harm that is inflicted on them. The dichotomous structure of a minimal goal on the other hand abets the internal attribution of a deviation from a superordinate goal (example: "The OG deviated from the average contribution goal of 80 % because they wanted to free-ride and profit from our high contributions. The goal was so simple you could only miss it if you wanted to!"). An internal attribution of a deviation is promoted by the simplicity of a minimal goal as it can only be fully achieved or not at all. Individuals or groups who deviate from a superordinate goal that is represented as a minimal goal cannot hope for an evaluation of their deviance based on a "benefit-of-the-doubt principle, simply because there is no doubt. This internal attribution in the case of a minimal goal representation not only inhibits an attenuation of the perceived magnitude of harm caused by an OG´s deviation. It rather extenuates the perceived magnitude of harm as the OG is seen as the sole reason for the damage done.

This suspected effect of goal-type on the attribution of deviance is hypothetical. It appears however highly plausible as it complements the effect of goal-type on the perceived magnitude of harm caused by the OGs deviance and the resulting differences in negative moral emotions toward the OG and the extent of harm that is inflicted on members of the OG.

5.5 Perspectives on the Infliction of Harm on Deviants

It has been argued that there are two fundamental perspectives on the infliction of harm on deviants. A first perspective has been focusing on the perspective of the targets of harmful behavior such as victims of discrimination or even open intergroup aggression as in the case of ethnic cleansings. From this perspective any factor that has the potential to increase the extent of harm inflicted on deviants - such as a minimal goal representation of a superordinate goal - has to be condemned. The question how the formation of a minimal goal representation of a superordinate goal could be averted appears self evident from a victim´s point of view. A second point of view however has taken a more inclusive and functional perspective on the infliction of harm on deviants. This approach takes the perspective of an inclusive group that is faced with the deviance of some of its members. This point of view is based on the

assumption that any group goal may be described on an abstract level as a Public Goods Game. The punishment of deviants enables the sustainment of a high contribution rate towards the Public Good of the group goal. This inclusive perspective suggests a completely different appraisal of the central finding that the deviation from a minimal superordinate goal leads to a higher level of harm inflicted on deviants than the deviation from a maximal superordinate goal. This higher level of harm inflicted on deviants should incentivize a persistently higher level of contributions to the Public Good of the group goal. The representation of a superordinate goal as a minimal goal should lead in other words to a more effective pursuit of any designated group goal.

The peril of an unconsidered application of this highly abstract interpretation to real circumstances lies at hand. The capability of a minimal superordinate goal to motivate group members to strive more ambitiously towards a common group goal might easily be misinterpreted or misused as a justification for any infliction of harm on deviants. Therefore any application of the functional implications of the goal-type of a superordinate goal has to be accompanied by a normative evaluation of the respective superordinate goal and the measures of harm involved. A minimal superordinate goal such as the ethnic and religious homogeneity of a population in a given region that has often been at the root of ethnic cleansings is for example strictly unacceptable from the widely acknowledged normative standpoint of human rights. The forms of harm inflicted on deviants in many of these cases such as murder and mutilation are equally unacceptable. However, the minimal goal that prisoners of war must not be tortured and humiliated is a minimal goal acknowledged in most western countries. This superordinate goal that no torture must be used against prisoners of war is a minimal goal as no deviation from it can be tolerated. Groups of individuals deviating from this goal - as it has been the case for example in Abu Ghuraib where Iraqi prisoners of war were tortured and abused - should be confronted with severe but measured consequences such as imprisonment and dishonorable discharge in order to ensure the upkeep of this important goal. It has been rightly argued that the punishment of deviants allows the upkeep of any group goal whatsoever (Boyd & Richerson, 1992). It has nevertheless to be kept in mind that not all group goals are equal from a normative standpoint. All group goals must be subjected to a normative ultima ratio such as the human rights as a minimal, utmost superordinate goal.

Recent research on cooperation in social dilemmas adds in addition a promising perspective to this inclusive, functional perspective on the infliction of harm on deviant. It has

been shown that not only punishment, but also the assignment of reward has the capability to elicit and maintain high levels of cooperation in Public Goods Games. It has been shown that a reward-option can be as successful as a punishment-option in order to achieve and sustain a high level of cooperation in a repeated Public Goods Game with individual identities persisting across rounds (Rand, Dreber, Ellingsen, Fudenberg, Nowak, 2009; see also: Sefton, Shupp, Walker, 2006). Similarly, Hauert and co-authors could show that it is the freedom to abstain from a group endeavor that enables the emergence of cooperation based on the punishment of defectors (Hauert, Traulsen, Brandt, Nowak, Sigmund, 2007). These results illustrate that the punishment of defectors is just one strategy to elicit and maintain cooperation in groups among others. Future research should consider therefore not only the effect of the goal-type of a superordinate goal on the infliction of harm on deviants. But also to inquire a potential effect of goal-type on the assignment of rewards to group members and the effect of rewards on the individual tendency to contribute to a group endeavor. The present research has shown that the deviation from a minimal superordinate goal leads to higher levels of harm inflicted on deviants from this goal than the deviation from a maximal goal. It has also been suggested that minimal and maximal goals serve different social functions. Minimal goals may define the borders of a group, making a dichotomous representation of a superordinate goal and the severe punishment of any deviants from this goal appropriate in order to protect the group´s cohesion and functionality. Maximal goals may define the prototype of a group, an ideal that group members should strive for. It seems plausible that not the infliction of harm on any deviants is the best suited measure to incentivize the group members´ striving towards such a maximal superordinate goal, but the assignment of rewards.

This doctoral thesis contributes in sum to the understanding of why and to what extent members of a deviant group are punished. Five studies were conducted to test the hypothesis that the deviation from a minimal superordinate goal leads to a higher level of harm inflicted on the members of a deviant group than the deviation from a maximal superordinate goal. This relation was expected to be mediated by negative moral emotions toward the deviant group. An experimental paradigm was specifically designed to test these hypotheses, combining a strictly experimental setup with the measurement of actual forms of harm inflicted on members of a deviant group. The evidence yielded from these studies supports both hypotheses with great consistency across all five studies and for several forms of harm. Various alternative explanations for this effect of the goal-type of a superordinate goal on the

infliction of harm on members of a deviant group could be ruled out by incremental changes in the experimental design across the course of the different studies.

The presented research provides therefore strong support for the importance of the goal-type of a superordinate goal for the infliction of harm on deviants, thus contributing to the understanding of factors determining the intentional maltreatment of groups of deviants.

6 Literature

Adair, J. G., Dushenko, T. W., & Lindsay, R. C. L. (1985). Ethical regulations and their impact on research practice. *American Psychologist, 40*, 59-72.

Allport, G. (1954). *The Nature of Prejudice*. Cambridge MA: Addison-Wesley.

Allison, P. D. (1992). The cultural evolution of beneficient norms. *Social Forces., 71*, 279-301.

Alexander, R. (1979). *Darwinism and Human Affairs.* Seattle: University of Washington Press.

American Psychological Association (APA) (2001). *The Publication Manual of the American Psychological Association*. 5th ed. Washington, DC: American Psychological Association.

Axelrod, R. (1984). *The Evolution of Cooperation.* New York: Basic Books.

Aronson, E., & Carlsmith, J. M. (1968). Experimentation in social psychology. In G. Lindzey & E. Aronson (Eds.), *The handbook of social psychology* (2nd ed.). Reading, MA: Addison-Wesley.

Austin, J. T., & Vancouver, J. B. (1996). Goal Constructs in Psychology: Structure, Process, and Content. *Psychological Bulletin, 120*, 338-375.

Axelrod, R. & Hamilton, W. D. (1981). The evolution of cooperation. *Science, 211*, 1390-1396.

Bandura, A., Underwood, B., & Fromson, M. E. (1975). Disinhibition of Aggression through Diffusion of Responsibility and Dehumanization of Victims. *Journal of Research in Personality, 9,* 253-269.

Bar-Tal, D. (1989). Delegitimization: The Extreme Case of Stereotyping and Prejudice. In: Bar-Tal, D., Grauman, C. F., Kruglanski, A. W., & Stroebe, W. (Eds.) Stereotyping and prejudice: Changing conceptions, 169-182, New York: Springer-Verlag.

Bar-Tal, D. (1990). Causes and consequences of deligitimization: Models of conflict and ethnocentrism. *Journal of Social Issues, 46,* 65-81.

Bar-Tal, D. (2004). The necessity of observing real life situations: Palestinian-Israeli violence as a laboratory for learning about social behaviour. *European Journal of Social Psychology, 34,* 677-701.

Baumeister, R. F., & DeWall, C., Ciarocco, N. J., & Twenge, J. M. (2005). Social Exclusion Impairs Self-Regulation. *Journal of Personality and Social Personality,88, 589-604.*

Baumeister, R. F., & Leary, M. R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin, 117*, 497-529.

Baumeister, R. F., Twenge, J. M., & Nuss, C. K. (2002). Effects of Social Exclusion on Cognitive Processes: Anticipated Aloneness Reduces Intelligent Thought. *Journal of Personality and Social Personality,* 83, 817-827.

Baumrind, D. (1964). Some Thoughts on Ethics of Research: After Reading Milgram´s ”Behavioral Study of Obedience”. *American Psychologist, 19*, 421-423.

Becker, H. S. (1963).*Outsiders: Studies in the Sociology of Deviance*. New York: The Free Press.

Berthold, A., Mummendey, A., Kessler, T., Lücke, B., & Schubert, T.(submitted). Different or worse? The influence of two different types of goals on ingroup projection and attitudes towards outgroups.

Blanz, M., Mummendey, A., & Otten, S. (1995a). Perceptions of relative group size and group status: Effects on intergroup discrimination in negative evaluations. *European Journal of Social Psychology, 25,* 231-247.

Blanz, M., Mummendey, A., & Otten, S. (1995b). Positive-negative asymmetry in social discrimination: The impact of stimulus-valence, size- and status-differentials in intergroup evaluations. *British Journal of Social Psychology, 34,* 409-419.

Bobo, L. (1983). White´s opposition to busing: Symbolic racism or realistic group conflict? *Journal of Personality and Social Psychology, 45,* 1196-1210.

Boyd, R., & Richerson, P. J. (1992). Punishment allows the evolution of cooperation (or anything else) in sizable groups. *Ethology and Sociobiology, 13,* 171-195.

Brewer, M. B. (1979). In-group bias in the minimal intergroup situation: A cognitive motivational analysis. *Psychological Bulleti*n, 86, 307-324.

Brewer, M. B., (1991). The social self: On being the same and different at the same time. Personality and Social Psychology Bulletin, 17, 475-482.

Brewer, M., B. (1999). The Psychology of Prejudice: Ingroup Love or Outgroup Hate? *Journal of Social Issues, 55,* 429-444.

Brewer, M. B. (2001). Ingroup identification and intergroup conflict: When does ingroup love become outgroup hate? In R. D. Ashmore, & L. Jussim (Eds.), *Social Identity, Intergroup Conflict, and Conflict Reduction*, New York: Oxford University Press.

Brewer, M.B. (1991). The social self: On being the same and different at the same time. *Personality and Social Psychology Bulletin, 17,* 475-482.

Brock, B., & Haslam, N. (2010). Excluded from humanity: The dehumanizing effects of social ostracism. *Journal of Experimental Social Psychology, 46,* 107-113.

Brown, R. J. (1995). *Group processes: Dynamics within and between groups* (2nd ed.). Oxford, UK: Blackwell.

Buhl, T. (1999). Positive-Negative Asymmetry in Social Discrimination: Meta-Analytical Evidence. *Group Processes and Intergroup Relations, 2,* 51-58.

Campbell, D. T. (1975). On the conflicts between biological and social evolution and between psychology and moral tradition. *American Psychologist, 30,* 1103-1126.

Caporael, L. R., Dawes, R. M., Orbell, J. M., & van de Kragt, A. J. C. (1989). Selfishness examined: Cooperation in the absence of egoistic incentives. *Behavioral and Brain Sciences, 12,* 683-739.

Carlsmith, Darley, & Robinson (2002). Why Do We Punish? Deterrence and Just Deserts as Motives for Punishment. *Journal of Personality and Social Psychology, 83*, 284-299.

Castano, E., Yzerbyt, V., Bourguignon, D., & Seron, E. (2002). Who May Enter? The Impact of In-Group Identification on In-Group/Out-Group Categorization. *Journal of Experimental Social Psychology, 38,* 315-322.

Chow, M. R., Tiedens, L. Z., & Govan, C. L. (2007). Excluded emotion: The role of anger in antisocial responses to ostracism, *Journal of Experimental Social Psychology, 44*, 896-903.

Christensen, L. (1988). Deception in Psychological Research: When Is Its Use Justified? *Personality and Social Psychology Bulletin, 14,* 664-675.

Cialdini, R., B., Goldstein, N., J., (2004). Social Influence: Compliance and Conformity, *Annual Review of Psychology, 55,* 591-621.

Cook, K. S., & Yamagishi, T. (2008). A Defense of Deception on Scientific Grounds. *Social Psychology Quarterly, 71,* 215-221.

Cialdini, R. B., Kallgren, C. A., & Reno, R. R. (1991). A focus theory of normative conduct: A theoretical refinement and re-evaluation of the role of norms in human behavior. *Advances in Experimental Social Psychology, 24*, 201-234.

Corneille, O., Yzerbyt, V. Y., Rogier, A., & Buidin, G. (2001). Threat and the Group attribution Error: When Threat Elicits judgements of Extremity and Homogeneity. *Personality and Social Psychology Bulletin, 27,* 437-446.

Cotrell, C. A., & Neuberg, S., L. (2005). Different Emotional Reactions to Different Groups: A Sociofunctional Threat-Based Approach to "Prejudice". *Journal of Personality and Social Personality, 88*, 770-798.

Darley, J. M., Carlsmith, K. M., & Robinson, P. H (2000). Incapacitation and Just Deserts as Motives for Punishment. *Law and Human Behavior, 24,* 2000.

Davis, D., D. & Holt, C. A. (1993). *Experimental economics.* Princeton University Press.

Dawes, R, & Thaler, R. (1988). Anomalies: Cooperation. *Journal of Economic Perspectives, 2,* 187-197.

DeWall, C. N., Twenge, J. M., Gitter, S. A., & Baumeister, R. F. (2009). It´s the Thought That Counts: The Role of Histile Cognition in Shaping Aggressive Responses to Social Exclusion. *Journal of Personality and Social Personality, 96,* 45-59.

Diehl, M. (1990). The Minimal Group Paradigm: Theoretical Explanations and Empirical Findings. In Stroebe, W. & Hewstone, M. (Eds.), European review of social Psychology (Vol. 1, pp. 263-292). New York:Wiley.

Dollard, J., Doob, L. W., Miller, N. E., Mowrer, O. H., & Sears, R. R. (1939). *Frustration and aggression*. New Haven, CT: Yale University Press.

Doosje, B., Spears, R., & Koomen, W. (1995). When bad Isn´t all Bad: Strategiec use of Sample Information in generalization and Stereotyping. *Journal of Personality and Social Personality, 69*, 642-655.

Eisenberger, N. I., Lieberman, M. D., & Williams, K. D. (2003). Does Rejection Hurt? An fMRI Study of Social Exclusion. *Science, 302*, 290-292.

Eisenberger, N. I., Jarcho, J., M., Liebermann, M. D., & Naliboff, B. D. (2006). An experimental study of shared sensitivity to physical pain and social rejection. *Pain, 126*, 132-138.

Ellsworth, P. C. & Ross, L. (1983). Public opinion and capital punishment: A close examination of the views of abolitionists and retentionists. Crime Delinquency, *29,* 116-169.

Esses, V. M., Dovidio, J. F., Jackson, L. M., & Armstrong, T. L. (2001). The immigration dilemma: The role of perceived group competition, ethnic prejudice, and national identity. *Journal of Social Issues, 57,* 389-412.

Esses, V. M., Haddock, G., & Zanna, M. P. (1993). Values, stereotypes, and emotions as determinants of intergroup attitudes. In D. M. Mackie & D. L. Hamilton (Eds.), Affect, cognition and stereotyping: Interactive processes in group perception (pp. 137-166). New York: Academic Press.

Esses, V. M., Jackson, L. M., & Armstrong, T. L. (1998). Intergroup Competition and Attitudes Toward Immigrants and Immigration: An Instrumental model of Group Conflict, *Journal of Social Issues, 54,* 699-724.

Fahrenberg, J., Hampel,R., & Selg, H.(2001): *Freiburger Persönlichkeitsinventar (FPI). Revidierte Fassung (FPI-R) und teilweise geänderte Fassung (FPI-A1)*. Hogrefe, Göttingen. (7.Edition).

Fehr, E., & Fischbacher, U. (2004a). Social norms and human cooperation. *Trends in Sciences, 8,* 185-190.

Fehr, E., & Fischbacher, U. (2004b). Third-party punishment and social norms. *Evolution and Human Behaviour, 25,* 63-87.

Fehr, E., & Gächter, S. (2000). Cooperation and punishment in public good experiments. *American Economic Review, 90*, 980-994.

Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature, 415*, 137-140.

Festinger, L (1954). A theory of social comparison processes. *Human Relations, 7,* 117-140.

Fischer, A. H., & Manstead, A.S.R. (2008). Social Functions of Emotions. In M. Lewis, J. M. Haviland-Jones, & L. Feldman Barrett (Eds.) *Handbook of Emotions.* (456-468). The Guilford Press, New York London.

Fischer, A. H., & Roseman, I. J. (2007). Beat Them or Ban Them: The Characteristics and Social Functions of Anger and Contempt. *Journal of Personality and Social Personality, 93,* 103-115.

Frank, S. A. (2003). Perspective: Repression of Competition and the Evolution of Cooperation, *Evolution, 57,* 693-705.

Fritsche, I., Kessler, T., Mummendey, A., & Neumann, J. (2009). Minimal and maximal goal orientation and reactions to norm violations. *European Journal of Social Psychology, 39,* 3-21.

Fritsche, I., & Schubert, T.W. (2009). Go to hell! Determinants of motivated social exclusion. In: Otten, S, Sassenberg, K., Kessler, T. (Eds.): *Intergroup Relations: The Role of Motivation and Emotion.* New York, NY, US: Psychology Press.

Galinsky, A. D., & Moskowitz, G. B. (2000). Perspective taking: Decreasing stereotype expression, stereotype accessibility and in-group favoritism. Journal of Personality and Social Psychology, 78, 708-724.

Gintis, H., Bowles, S., Boyd, R., & Fehr, E. (2003). Explaining altruistic behavior in humans. *Evolution and Human Behavior, 24,* 153-172.

Gollwitzer, P. M. (1990). Action phases and mind-sets. In E. T. Higgins & R. M. Sorrentino (Eds.), *The handbook of motivation and cognition: Foundations of social behavior* (Vol. 2, pp. 53-92). New York: Guilford Press.

Gonsakorale, K., & Williams, K. D. (2007). The KKK won´t let me play: Ostracism even by a despised outgroup hurts. European *Journal of Social Psychology, 37,* 1176-1186.

Gross, A. E., & Fleming, I. (1982). Twenty years of deception in social psychology. *Personality and Social Psychology Bulletin, 8* , 402-408.

Hamilton, D. W. (1964). The genetical evolution of social behavior. *Journal of Theoretical Biology,* 7, 1-52.

Hardy, C., & van Vugt, M. (2006). Nice guys finish first: The competitive altruism hypothesis. *Personality and Social Psychology Bulletin, 32,* 1402-1423.

Haslam, N. (2006). Dehumanization: An intergrative review. *Personality and Social Psychology Review, 10,* 252-264.

Hauert, C., Traulsen, A., Brandt, H., Nowak, M., & Sigmund, K. (2007). Via freedom to coercion: the emergence of costly punishment, *Science, 316,* 1905-1907.

Henderson-King, E. I. & Nisbett, R. E. (1996). Anto-Black Prejudice as a Function of Exposure to the Negative Behavior of a Single Black Person. *Journal of Personality and Social Personality, 71,* 654-664.

Henrich, J., & McElreath, R. (2003). The Evolution of Cultural Evolution. *Evolutionary Anthropology, 12,* 123-135.

Hertwig, R., & Ortmann, A. (2001). Experimental practices in economics: A methodological challenge for psychologists? *Behavioral and Brain Sciences, 24*, 282-451.

Hey**,** J.D.**, (**1991**),** *Experiments in Economics*, Oxford**:** Blackwell.

Horwitz, M., & Rabbie, J.M. (1982). Individuality and membership in the intergroup system. In: H. Tajfel (Ed.), Social Identity and Intergroup relations, 241-274, London: Academic Press.

Kelman, H. C. (1967). Human use of human subjects: The problem of deception in social psychological experiments. *Psychological Bulletin, 67*, 1-11.

Kenrick, D. T., Maner, J. K., Butner, J., Li, N. P., Becker, D. V., & Schaller, M. (2002). Dynamical Evolutionary Psychology: Mapping the Domains of the New Interactionist Paradigm, *Personality and Social Psychology Review, 6,* 347-356.

Kerr, N., L., & Levine, J. M. (2008). The Detection of Social Exclusion: Evolution and Beyond. *Group Dynamics: Theory, Research, and Practice, 12,* 39-52.

Kessler, T., Neumann, J., Mummendey, A., Berthold, A., Schubert, T. & Waldzus, S. (2010). How do we assign punishment? The impact of minimal and maximal standards on the evaluation of deviants. *Personality and Social Psychology Bulletin, 36,* 1213-1224.

Kimmel**,** A. J. ***(***1996***).*** *Ethical issues in behavioral research: A survey***.** Cambridge**,** MA**:** Blackwell**.**

Kinder, D. R., & Sears, D. O. (1981). Prejudice and politics: Symbolic racism versus racial threats to the good life. *Journal of Personality and Social Psychology, 40,* 414-431.

Kling, A., Lancaster, J., & Benitone, J. (1970). Amygdalectomy in the free-ranging vervet (Cercooithecus aethiops). *Journal of Psychiatric Research*, 7, 191-199.

Kollock, P. (1998). Social Dilemmas: The Anatomy of Cooperation. *Annual Review Sociolology, 24,* 183-214.

Krebs, J. R., Stephens, D. W., & Sutherland, W. J. (1983). Perspectives in foraging theory. in. Brush & G. Clark Jr. (Eds.), *Perspectives in ornithology* (pp. 165-221). New York: Cambridge University Press.

Kurzban, R., & Leary, M. R. (2001). Evolutionary Origins of Stigmatization: The Functions of Social Exclusion. *Psychological Bulletin, 127*, 187-208.

Ledyard, J. (1995). Public goods. A survey of experimental research. In: *Handbook of Experimental Economics*, Kagel, J. and Roth, A, (Eds.), 111-194, Princeton University Press.

Levin, S., & Sidanius, J. (1999). Social dominance and social identity in the United States and Israel: Ingroup favoritism or outgroup derogation? *Political Psychology, 20*, 99-126.

Leyens, J.-P., & Yzerbyt, V. Y. (1992). The ingroup overexclusion effect: Impact of valence and confirmation on stereotypical information search. *European Journal of Social Psychology, 22,* 549-569.

MacDonald, G., & Leary, M. R. (2005). Why Does Social Exclusion Hurt? The Relationship Between Social and Physical Pain. *Psychological Bulletin, 131,* 202-223.

Maslow, A. H. (1968). *Toward a psychology of being.* New York: Van Nostrand.

Messick, D. W., & Mackie, D. M. (1989). Intergroup relations. *Annual Review of Psychology,* 40, 45-81.

Miller, N., Pedersen, W. C., Earlywine, M., & Pollock, V. E. (2003). A Theoretical Model of Triggered Displaced Aggression. *Personality and Social Psychology Review, 7,* 75-97.

Mummendey, A., & Otten, S. (1998). Positive-Negative Asymmetry in Social Discrimination. *European Review of Social Psychology, 9,* 107-143.

Mummendey, A., & Otten, S., Berger, U., & Kessler, T. (2000). Positive-negative Asymmetry in Social Discrimination: Valence of Evaluation and Salience of Categorization, *Personality and Social Psychology Bulletin, 26,* 1258-1270.

Mummendey, A., Otten, S., Berger, U., & Kessler, T. (2000). Positive-negative asymmetry in social discrimination: Valence of evaluation and salience of categorization. *Personality and Social Psychology Bulletin, 26,* 1258-1270.

Mummendey, A., & Simon, B. (1991) Diskriminierung von Fremdgruppen: zur Asymmetrie im Umgang mit positiven und negativen Bewertungen und Ressourcen. In D. Frey (Ed.), Berichte über den 37. Kongress der Deutschen Gesellschaft für Psychologie in Kiel, 1990, 2, 359-365. Göttingen: Hogrefe.

Mummendey A., Simon, B., Dietze C., Grünert, M., Haeger, G., Kessler, S., Lettgen S., & Schäferhoff, S. (1992). Categorization is not enough: Intergorup discrimination in negative outcome allocation, *Journal of Experimental Social Psychology, 28,* 125-144.

Mummendey, A., Waldzus, S., & Kessler. T. (2002). Psychologische Grenzen von Toleranz. Voraussetzungen und Determinanten expliziter Ablehnung und Ausgrenzung von Fremdgruppen. Jena: unpublished grant proposal.

Mummendey, A., & Wenzel, M. (1999). Social discrimination and tolerance in intergroup relations: Reactions to intergroup difference. *Personality and Social Psychology Review, 3*, 158-174.

Nowak, A., & Vallacher, R. R. (1998). *Dynamical social psychology*. The Guilford, Press New York.

Opotow, S. (1990). Moral Exclusion and Injustice: An Introduction. *Journal of Social Issues, 46,* 1, 1-20.

Opotow, S. (1995). Drawing the line: Social categorization, moral exclusion, and the scope of justice. In B. B. Bunker & J. Rubin (Eds.), *Conflict, cooperation and justice: Essays inspired by the work of Morton Deutsch* (pp. 347-369). San Francisco: Jossey-Bass.

Ostrom, T.M., & Sedikides, C. (1992). Out-Group Homogeneity Effects in Natural and Minimal Groups. *Psychological Bulletin, 112*, 536-552.

Ostrom, E. (1992). Crafting Institutions for Self-Gover ning Irrigation Systems. San Francisco: Institute for Contemporary Studies.

Ostrom, E. (2000). Collective action and the evolution of social norms. *Journal of Economic Perspectives, 14,* 137-158.

Otten, S., Mummendey, A., & Blanz, M. (1996) . Intergroup discrimination in positive and negative outcome-allocations: The impact of stimulus-valence, relative group status and relative group size. *Personality and Social Psychology Bulletin, 22,* 568-581.

Nowak, A., & Vallacher, R. R. (1998). Dynamical Social Psychology: New York: Guilford.

Panksepp, J. (1998). *Affective Neuroscience: The Foundations of Human and Animal Emotions.* New York: Oxford University Press

Pettigrew, T. F., Allport, G. W., & Barnett, E.O. (1958). Binocular resolution and perception of race in South Africa. British Journal of Psychology, 49, 265-278.

Price, M. E., Cosmides, L. & *Tooby*, J. (2002). Punitive sentiment as an anti-free rider psychological device. *Evolution and Human Behavior, 23,* 203-231.

Pruitt, D. G., & Kimmel, M. J. (1977). Twenty years of experimental gaming: Critique, synthesis, and suggestions for the future. *Annual Review of Psychology, 28*, 363-392.

Rabbie, J. M., Horwitz, M. (1969). Arousal of ingroup-outgroup bias by a chance win or loss. *Journal of Personality and Social Psychology, 13,* 269-277.

Rand, D. G., Dreber, A., Ellingsen, T., Fudenberg, D. & Nowak, M. A. (2009). Positive interactions promote public cooperation, *Science, 325,* 1272-1275.

Rapoport, A., (1985). Provision of public goods and the MCS experimental paradigm. American Political Science Review *79***,** 148-155.

Rapoport, A., (1987). Research paradigms and expected utility models for the provision of step-level public goods. Psychological Review, *94***,** 74-83.

Riek, B. M., Mania, E. W., & Gaertner, S. L. (2006). Intergroup Threat and Outgroup Attitudes: A Meta-analytic Review. *Personality and Social Psychology Review, 10,* 336-353.

Rozin, P., Lowery, L., Imada, S., & Haidt, J. (1999). The CAD Triad Hypothesis: *Journal of Personality and Social Psychology, 76,* 574-586.

Sassenberg, K., Kessler, T., & Mummendey, A. (2003). Less negative = more positive? Social discrimination as avoidance or approach. *Journal of Experimental Social Psychology, 39,* 48-58.

Schaller, M., & Latané, B. (1996). Dynamic social impact and the evolution of social representations: A natural history of stereotypes. *Journal of Communication, 46,* 64-71.

Sears, D. O. (1988). Symbolic racism. In P. A. Katz & D. A. Taylor (Eds*.), Eliminating racism: Profiles in controversy* (pp. 53-84). New York: Plenum Press.

Sefton, M., Shupp, R., & Walker, J. (2006). The Effect of Rewards and Sanctions in Provision of Public Goods. *Center For Applied Economics and Policy Research Working Paper*, Bloomington.

Sherif, M. (1936). *The psychology of social norms*. New York: Harper.

Sherif, M., White, B. J., & Harvey, O.J. (1955). Status in experimentally produced groups, *American Journal of Sociology, 60,* 370-79.

Sherif, M., Harvey, O. J., Hood, W. R., & Sherif, C. W. (1961). Intergroup Conflict and Cooperation: The Robbers Cave Experiment. Norman, OK: University Book Exchange.

Sherif, M. (1966) I*n common predicament: Social psychology of intergroup conflict and cooperation*, Boston: Houghton-Mifflin.

Silk, J. B., Alberts, S. C.. & Altmann, J. (2003). *Social bonds of female baboons enhance infant survival. Science, 302,* 1231-1234.

Smith, S. S., & Richardson, D. (1983). Amelioration of deception and harm in psychological research: The important role of debriefing. *Journal of Personality and Social Psychology, 44*, 1075-1082.

Staub, E. (1989). *The roots of evil: origins of genocide and other group violence*. New York: Cambridge University Press.

Staub**,** E. (1999***).*** The roots of evil***:*** Personality**,** social conditions*,* culture and basic human needs. *Personality and Social Psychology Review, 3,* 179-192.

Stephan, W.G., & Stephan, C. W. (1996). Predicting Prejudice. *International Journal of Intercultural Relations, 20,* 409-426.

Stephan, W.G., & Stephan, C. W. (2000). *An integrated theory of prejudice.* In. Oskamp (Ed.), Reducing prejudice and discrimination: The Claremont Symposium on applied social psychology (pp. 23-45). Mahwah, NJ: Lawrence Erlbaum Associates, inc.

Stillman, T. F., Baumeister, R. F., Lambert, N. M., Crescioni, A., W., DeWall, C. N., & Fincham, F. D. (2009). Alone and without purpose: Life loses meaning following social exclusion. *Journal of Experimental Social Psychology, 45*, 686-694.

Struch, N., & Schwarz, S. H. (1989). Intergroup aggression: Its predictors and distinctness from in-group bias. *Journal of Personality and Social Psychology, 56*, 364-373.

Sumner, W. G. (1906). *Folkways.* New York: Ginn.

Tajfel, H., Billig, M.G., Bundy, R.P., & Flament, C. (1971). Social categorization and intergroup behaviour. *European Journal of Social Psychology, 1,* 149-178.

Tajfel, H. & Turner, J. C. (1979). An integrative theory of intergroup behaviour. In W. G. Austin & S. Worchel (Eds), *The Social Psychology of Intergroup Relations*, 33-47, Monterey, CA: Brooks/Cole.

Tajfel, H. & Turner, J. C. (1986). The social identity theory of inter-group behavior. In S. Worchel & L. W. Austin (Eds*.), Psychology of Intergroup Relations*. Chigago: Nelson-Hall

Trivers, R. L. (1971). The evolution of reciprocal altruism. *Quarterly Review of Biology, 46,* 35-57.

Turner, J. C., Hogg, M. A., Oakes, P. J., Reicher, S. D., & Wetherell, M. S. (1987). *Rediscovering the social group. A self-categorization theory.* Oxford: Basil Blackwell.

Twenge, J. M., Baumeister, R. F., Tice, & Stucke (2001). If You Can´t Join Them: Effects of Social Exclusion on Aggressive Behavior. *Journal of Personality and Social Psychology, 81,* 1058-1069.

Twenge, J. M., Catanese, K. R., & Baumeister, R. F. (2002). Social Exclusion causes Self-Defeating Behavior. *Journal of Personality and Social Psychology, 83,* 606-615.

Twenge, J. M., Catanese, K. R., & Baumeister, R. F. (2003). Social Exclusion and the Deconstructed State: Time Perception, Meaninglessness, Lethargy, Lack of Emotion, and Self-Awareness, *Journal of Personality and Social Psychology, 85,* 409-423.

Twenge, J. M., Baumeister, R. F., DeWall, C. N., Ciarocco, N. J., & Bartels, J. M. (2007). Social Exclusion Decreases Prosocial Behavior. *Journal of Personality and Social Psychology, 92,* 56-66.

Tybur, J. M., Lieberman, D. & Griskevicius, V. (2009). Microbes, Mating, and Morality: Individual Differences in Three Functional Domains of Disgust. *Journal of Personality and Social Psychology, 97,* 103-122.

Vallacher, R., Coleman, P., Nowak, A., & Bui-Wrzosinska, L. (2010). Rethinking intractable conflict: The perspective of dynamical systems. *American Psychologist, 65,* 262-278.

Vallacher, R. R., Read, S. J., & Nowak, A. (2002). The dynamical perspective in personality and social psychology. *Personality and Social Psychology Review, 6,* 264-273.

Van Beest, I., & Williams, K., D. (2006). When Inclusion Costs and Ostracism Pays, Ostracism Still Hurts. *Journal of Personality and Social Psychology, 91,* 918-928.

Vanbeselaere, N. (1991). The impact of in-group and out-group homogeneity/heterogeneity upon intergroup relations. *Basic and Applied Social Psychology, 12 ,* 291-301.

Van de Kragt, A. J. C. *,* Orbell, J. M.*, &* Dawes, R. M. *(*1983*).* The minimal contributing set as a solution to public goods problems*. American Political Science Review,* 77*,* 112-122.

Van Kleef, G. A., & Cote, S. (2007). *Journal of Applied Psychology, 92,* 1557-1569.

Van Kleef, G. A., & De Dreu, C. K. W., & Manstead, A. S. R. (2004). The Interpersonal Effects of Emotions in Negotiations: A Motivated Information Processing Approach. *Journal of Personality and Social Psychology, 87*, 510-528.

Vasquez, E. A., Lickel, B., Hennigan, K. (2010). Gangs, displaced, and group-based aggression. *Aggression and Violent Behavior, 15,* 130-140.

Vidmar, N. (1974). Retributive and utilitarian motives and other correlates of Canadian attitudes toward the death penalty. Canadian Psychology, ***15***, 337-356.

Vidmar, N., & Ellsworth, P. (1974). Public opinion and the death penalty. Stanford Law Review*, 26*, 1245-1270.

Waldzus, S., Mummendey, A., Wenzel, M., & Weber, U., (2003). Towards tolerance: Representations of superordinate categories and perceived ingroup prototypicality. *Journal of Experimental Social psychology, 39,* 31-47.

Waldzus, S., Mummendey, A., & Wenzel, M. (2005). When „different“ means „worse“: In-group prototypicality in changing intergroup contexts. *Journal of Experimental Social Psychology, 41,* 76-83.

Warburton, W. A., Williams, K. D., & Cairns, D. R. (2006). When ostracism leads to aggression: The moderating effects of control deprivation. *Journal of Experimental Social Psychology, 42*, 213-220.

Williams, K. D. (2007). Ostracism, *Annual Review of Psychology, 58,* 425-452.

Williams, K. D. (1997). Social ostracism. In R. M. Kowalski (Ed.), *Aversive interpersonal behaviors* (pp. 133-170). New York: Plenum.

Williams, K. D., Cheung, C. K. T., & Choi, W. (2000). Cyberostracism: Effects of being ignored over the Internet. *Journal of Personality and Social Psychology, 79,* 748-762.

Williams, K. D, & Jarvis, B. (2006). Cyberball: A program for use in research on interpersonal ostracism and acceptance, *Behavior Research Methods, 38,* 174-180.

Wit, A. P., & Kerr, N. L. (2002). „Me Versus Just Us Versus Us All” Categorization and Cooperation in Nested Social Dilemmas. *Journal of Personality and Social Psychology, 83,* 616-637.

Yamagishi, T. (1986). The provision of a sanctioning system as a public good, *Journal of Personality and Social Psychology, 51,* 110-116.

Yamagishi, T. (1988). Seriousness of Social Dilemmas and the provision of a sanctioning system. *Social Psychology Quarterly, 51,* 32-42.

Yzerbyt, V. Y., & Leyens, J.-P., & Bellour, F. (1995). The ingroup overexclusion effect: identity concerns in decisions about group membership. *European Journal of Social Psychology, 25,* 1-16.

Zhong, C.-B., & Leonardelli, G. J. (2008). Cold and Lonely: Does Social Exclusion Literally Feel Cold? *Psychological Science, 19,* 838-842.

Zimring, F. E., & Hawkins, G. (1995). *Incapacitation*. New York: Oxford University Press.

Appendix

Group Categorization

1. Ich bin ungern mit Menschen zusammen, die ich noch nicht kenne.

2. Meine Bekannten halten mich für einen energischen Menschen.

3. Ich habe manchmal das Gefühl der Teilnahmslosigkeit und inneren Leere.

4. Es macht mir Spaß mit anderen zu wetteifern.

5. Ich bin auch an Wochenenden stark eingespannt.

6. Ich bin im Grunde ein eher ängstlicher Mensch.

7. Ich bin ziemlich lebhaft.

8. Ich schließe nur langsam Freundschaften.

9. Oft rege ich mich zu rasch über jemanden auf.

10. Meine Laune wechselt ziemlich oft.

Summary

This dissertation is concerned with determinants of explicit negative behavior towards members of groups that deviate from the goal of a superordinate group (SOG). It could be shown across five studies that the deviation from a minimal superordinate goal leads to a higher level of harm inflicted on members of a deviant OG than the deviation from a maximal superordinate goal (*goal-type hypothesis*). This effect of goal-type could even be found for the generalized social exclusion of uninvolved OG-members. It could be shown that the relation between goal-type (minimal/ maximal) of the superordinate goal and the level of harm inflicted on the deviants is mediated by negative moral emotions toward the OG (*negative emotions hypothesis*).

The presented research comprises two central aims. The first aim was a test of the assumption that the representation of a superordinate group goal as either minimal or maximal had a significant impact on the extent of harm that would be inflicted on members of a group deviating from this goal. Maximal goals are desired or undesired end states that can be approached or avoided. Events, which are evaluated relative to maximal goals are judged by their relative distance from these goals and therefore receive a graded evaluation. The valence of an event relative to a maximal goal depends on its distance to the goal. In contrast to this "shades of grey" type of evaluation, minimal goals lead to "black/ white" evaluations of an event. Minimal goals define a certain point that must be exceeded. A minimal goal can therefore only be fully met or not met at all. An event, which is evaluated relative to a minimal goal receives a dichotomous evaluation and is therefore either entirely positive or negative. It was assumed that the representation of a superordinate goal as either minimal or maximal not only affects the evaluation of an event relative to this goal, but also has strong implications on the behavioral reactions towards members of an OG that deviating from this superordinate goal. The second central aim of this dissertation was the design an experimental paradigm to test these hypotheses. This experimental paradigm addresses a number of methodological desiderata: First it allowed the introduction of the *superordinate goal* of making money by collectively contributing in a Public Goods Game. This superordinate goal was experimentally manipulated to resemble either a minimal or a maximal goal. Secondly *an intergroup context* was introduced on the basis of the Minimal Group Paradigm in order to allow the *deviation of a subgroup from the superordinate goal*. This categorization had no functional implications in terms of the Public Goods Game. The experimental paradigm allowed the measurement of actual *harm inflicted on the deviant subgroup* in response to the

deviation of this subgroup from the superordinate goal in the form of monetary punishment, social exclusion and generalized social exclusion. The entire experimental paradigm was a strictly *experimental setup* in order to exclude the influence of uncontrolled factors.

Five Studies were conducted in order to test both the *goal-type hypothesis* as well as the *negative emotions hypothesis*. The *superordinate goal* of making money by contributing collectively to a Public Goods Game was identical across all studies. In the first three studies the *goal-type manipulation* was based on slightly different designs of a Public Good Game. A Graded Public Goods Game was used to resemble the graded structure of a maximal superordinate goal. The superordinate goal consisted in this maximal goal condition in an ideal average contribution rate of 100 % the initial endowment resulting in a maximum payoff on a collective level. A step-level Public Goods Game was used to resemble the dichotomous "black/white" structure of a minimal goal. A threshold of an average contribution level had to be reached in order to receive a payoff from the Public Good.

An initial study provided support for the goal-type hypothesis: Participants in the minimal goal condition were more likely to punish members of the deviant OG monetarily. However, this relation between goal-type and monetary punishment was mediated by participants´ identification with their IG and not by negative moral emotions toward the OG. This mediation by identification was not hypothesized for. A second study was conducted to replicate the effect of goal-type on the infliction of harm on members of a deviant OG. This second study used a graded contribution option for the Public Goods Game. The effect of goal-type on the monetary punishment of members of a deviant OG that was found in the first study could be replicated. Participants in the minimal goal condition were more inclined to punish members of a deviant OG than participants in the maximal goal condition. This effect was as hypothesized mediated by negative emotions towards the OG, no by the participants´ identification with their IG. A third study added the dependent measures of social exclusion from future rounds and the generalized social exclusion of uninvolved OG-members from future profitable studies. In addition the goal-type manipulation was improved. Participants received in Study 2 - as a consequence of the OG´s deviation - a slightly lower payoff from the step-level Public Goods Game (minimal goal condition) than from the graded Public Goods Game (maximal goal condition). This difference in payoffs did not affect the main effect of goal-type on monetary punishment. Study 3 controlled experimentally for this difference in payoffs. The effect of goal-type on the infliction of harm on members of the deviant OG was found for all three measures of harm (monetary punishment, social exclusion

from future rounds and generalized social exclusion of uninvolved OG-members). This effect was mediated by negative moral emotions toward the OG for all three measures of harm. However, the structural differences between step-level Public Goods Game and graded Public Goods Game remained. In the maximal goal condition players received at least some payoff from the Public Good. In the minimal goal condition participants received no payment at all from the Public Good because of the deviation of the OG. This could have lead theoretically to a perception of the OG´s deviation as more severe, even though there were no differences in final payoffs in Study 3. Therefore a second manipulation of goal-type was introduced in Study 4, based on a mere framing of goal-type. All players engaged in a graded Public Goods Game. The superordinate goal was either framed as a graded "shades-of-grey"-goal or a minimal "black-or-white"-goal. The same effect tog goal-type as in the previous studies was found. Participants in the minimal goal condition were more likely to punish members of the deviant OG monetarily and exclude them from future rounds. This effect was again mediated by negative moral emotions towards the OG. A fifth study using the same manipulation of goal-type based on a mere framing as in Study 4 found could replicate the effect of goal-type on the generalized social exclusion of uninvolved OG-members from future profitable studies. A summarizing analysis could show the high stability of described effects across the various studies.

This doctoral thesis contributes in sum to an improved the understanding why members of deviant groups are sometimes punished mildly and sometimes confronted with the infliction of severe harm for the very same deviation. An ambitious experimental paradigm was designed that allows research on determinants of the infliction of harm on members of OG´s using actual behavioral measures of different forms of harm in an experimentally controlled context. Both initial hypotheses - *goal-type hypothesis* as well as *negative emotions hypothesis* - could be confirmed across multiple studies: Individuals with a minimal goal representation of a superordinate goal are more likely to inflict more harm on members of an OG deviating from this goal than individuals with a maximal goal representation. The presented research constitutes the introduction of the goal-type of a superordinate goal as an important determinant of the infliction of harm on members of deviant groups.

Zusammenfassung

Das zentrale Anliegen dieser Dissertation ist die Bestimmung von Determinanten explizit negativen Verhaltens gegenüber Mitgliedern von Fremdgruppen, die von einem Ziel einer übergeordneten Gruppe abweichen. Über 5 Studien hinweg konnte gezeigt werden, dass die Abweichung von einem als Minimalziel repräsentierten übergeordneten Ziel gegenüber den Mitgliedern einer abweichenden Gruppe zu einem höheren Grad an Bestrafung führt als die Abweichung von einem als Maximalziel repräsentierten übergeordneten Ziel (*Zieltyp-Hypothese*). Dieser Effekt des Zieltyps eines übergeordneten Ziels auf die Bestrafung von Mitgliedern einer devianten Gruppe konnte für verschiedene Formen von explizit negativem Verhalten wie monetäre Bestrafung, sozialen Ausschluß von zukünftigen Runden eines Public Good Games und sogar den generalisierten sozialen Ausschluß unbeteiligter Gruppenmitglieder gezeigt werden. Der Effekt des Zieltyps des übergeordneten Ziels (Minimalziel versus Maximalziel) auf die Bestrafung von Mitgliedern einer devianten Fremdgruppe wurde durch negative moralische Emotionen gegenüber der Fremdgruppe mediiert (*Negative Emotionen Hypothese*).

Die vorgestellte Forschungsarbeit umfasste zwei zentrale Zielsetzungen: Die erste Zielsetzung war der Test der Vermutung, dass die Repräsentation eines Ziels einer übergeordneten Gruppe als Minimal- oder Maximalziel einen entscheidenden Einfluss auf das Ausmaß der Bestrafung von Mitgliedern einer Gruppe hat, die von diesem übergeordneten Ziel abweichen. Ereignisse, die in Relation zu einem Maximalziel beurteilt werden, werden auf der Basis ihrer relativen Entfernung von diesem Ziel evaluiert. Je näher ein Ereignis einem Maximalziel kommt, desto positiver ist seine Bewertung. Im Gegensatz zu dieser Evaluation auf der Basis von “Grauschattierungen” erfolgt die Evaluation von Ereignissen in Relation zu Minimalzielen anhand eines dichotomen “schwarz/weiß”-Denkens. Minimalziele definieren eine bestimmte Schwelle, die nicht überschritten werden darf. Ein Minimalziel kann deshalb nur vollständig erfüllt werden oder überhaupt nicht. Ein Ereignis, das relativ zu einem Minimalziel evaluiert wird erfährt eine dichotome Bewertung und ist entweder vollständig positiv oder komplett negativ. Es wurde angenommen, daß die Repräsentation eines übergeordneten Ziels als Minimalziel oder Maximalziel nicht nur die Evaluation eines Ereignisses in Relation zu diesem Ziel beeinflußt, sondern auch weitreichende Implikationen für die Verhaltensreaktionen gegenüber den Mitgliedern einer Gruppe hat, die von diesem übergeordneten Ziel abweicht.

Die zweite zentrale Zielsetzung dieser Dissretation war die Gestaltung eines experimentellen Paradigmas, um sowohl *Zieltyp-Hypothese* als auch *Negative-Emotionen-Hypothese* zu testen. Dieses experimentelle Paradigma berücksichtigt eine Reihe von Desiderata: Es erlaubt die Einführung eines *gemeinsamen übergeordneten Ziels* in Form von Geld verdienen durch kollektives Beitragen in einem Public Goods Game. Dieses übergeordnete Ziel wurde experimentell manipuliert um entweder ein Minimalziel oder ein Maximalziel darzustellen (*Zieltypmanipulation*). Zudem wurde ein *Intergruppenkontext* auf der Basis des Minimalen Gruppenparadigmas eingeführt, um die Abweichung von Mitgliedern einer devianten Gruppe vorzubereiten. Diese willkürliche Kategorisierung in zwei Gruppen hatte keine funktionale Bedeutung für das Public Goods Game. Das experimentelle Paradigma erlaubte zudem den *Schaden zu messen*, welcher der vom gemeinsamen Ziel abweichenden Gruppe zugefügt wurde. Hierzu wurden unterschiedliche Maße wie monetäre Bestrafung, sozialer Ausschluß und generalisierter sozialer Ausschluß verwendet. Das gesamte Paradigma folgte einem *strikten experimentellen Aufbau* um den Einfluss unkontrollierter Faktoren auszuschließen.

Es wurden fünf Studien durchgeführt um sowohl *Zieltyp-Hypothese* als auch die *Negative Emotionen-Hypothese* zu testen. Das übergeordnete Ziel war in allen fünf Studien das Verdienen von Geld durch kollektives Beitragen in einem Public Goods Game. In den ersten drei Studien wurde der Zieltyp dieses übergeordneten Ziels (Minimalziel versus Maximalziel) durch Varianten eines Public Goods Games manipuliert. Ein kontinuierliches Public Goods Game wurde genutzt, um die stufenlose Struktur eines Maximalziels widerzuspiegeln. Das Maximalziel bestand in dieser Maximalzielbedingung in einer idealen durchschnittlichen Beitragsrate von 100 % der Anfangsauszahlungen, was aus kollektiver Sicht eine maximale Auszahlung aus dem Public Good zur Folge gehabt hätte. Ein gestuftes Public Goods Game wurde hingegen genutzt um die dichotome „schwarz/weiß“-Struktur eines Minimalziels widerzugeben. Hierbei musste eine bestimmte durchschnittliche Beitragsschwelle erreicht werden, damit alle Spieler eine Auszahlung aus dem Public Good erhielten.

Die Ergebnisse einer ersten Studie waren in Einklang mit der *Zieltyp-Hypothese*: Probanden in der Minimalzielbedingung bestraften Mitglieder einer devianten Gruppe stärker monetär als Probanden in der Maximalzielbedingung. Diese Beziehung zwischen dem Zieltyp des übergeordneten Ziels und monetärer Bestrafung wurde jedoch durch die Identifikation der Teilnehmer mit ihrer Eigengruppe mediiert, und nicht durch negative moralische Emotionen gegenüber der devianten Gruppe. Diese Mediation war somit unerwartet und widersprach der *Negative Emotionen Hypothese*. Eine zweite Studie wurde durchgeführt um den Effekt des

Zieltyps des übergeordneten Ziels auf das Ausmaß der monetären Bestrafung von Mitgliedern der devianten Fremdgruppe zu replizieren. In dieser zweiten Studie wurde im Rahmen des Public Goods Games eine stufenlose Beitragsoption verwendet. Der Effekt des Zieltyps des übergeordneten Ziels auf die monetäre Bestrafung der Mitglieder der devianten Gruppe konnte repliziert werden. Studienteilnehmer in der Minimalzielbedingung bestraften die Mitglieder der devianten Gruppe signifikant stärker als Probanden in der Maximalzielbedingung. Dieser Effekt wurde gemäß der *Negative-Emotionen Hypothese* durch negative moralische Emotionen mediiert, und nicht durch die Identifikation der Probanden mit ihrer Eigengruppe. In einer dritten Studie wurden zwei weitere abhängige Maße explizit negativen Verhaltens eingeführt: Der soziale Ausschluß von zukünftigen Runden des Public Goods Games und der generalisierte soziale Ausschluß von unbeteiligten Mitgliedern der devianten Gruppe. Zusätzlich wurde die Zieltypmanipulation weiter verbessert. Die Teilnehmer der zweiten Studie erhielten aufgrund der Abweichung der devianten Gruppe eine geringfügig geringere Auszahlung aus dem gestuften Public Goods Game (Minimalzielbedingung) als aus dem kontinuierlichen Public Goods Game (Maximalzielbedingung). Dieser geringe Unterschied in den Auszahlungen konnte jedoch nicht den Effekt des Zieltyps des übergeordneten Ziels auf das Ausmaß der monetären Bestrafung erklären. Dennoch wurden in der dritten Studie jegliche aus der Abweichung der devianten Gruppe resultierenden Auszahlungsunterschiede experimentell ausgeschlossen. Ein Effekt des Zieltyps des übergeordneten Ziels gemäß der Zieltyp-Hypothese wurde für alle drei abhängigen Maße gefunden (monetäre Bestrafung, sozialen Ausschluß von zukünftigen Runden und generalisierten sozialen Ausschluß unbeteiligter Mitglieder der devianten Gruppe). Dieser Effekt wurde bei allen drei abhängigen Maßen durch negative moralische Emotionen mediiert. Ein struktureller Unterschied zwischen gestuftem und kontinuierlichem Public Goods Game blieb auch bei der in Studie 3 verwendeten Manipulation des Zieltyps des übergeordneten Ziels übrig. Probanden in der Maximalzielbedingung erhielten trotz der Abweichung der devianten Gruppe eine geringe Auszahlung aus dem Public Good, wohingegen Probanden in der Minimalzielbedingung keinerlei Auszahlung aus dem Public Good erhielten. Obwohl es in der dritten Studie keinerlei Auszahlungsunterschiede zwischen Minimal- und Maximalzielbedingung gab, könnte dieser strukturelle Unterschied zu einer Wahrnehmung der Abweichung der devianten Gruppe als schwerwiegender in der Minimal- als in der Maximalzielbedingung geführt haben. Daher wurde in Studie 4 eine zweite Art der Zieltypmanipulation eingeführt. Diese basierte auf einem einfachen Framing des Zieltyps des übergeordneten Ziels als Minimal oder-Maximalziel. Alle Studienteilnehmer spielten ein

kontinuierliches Public Goods Game unabhängig von der Zeltypbedingung. Das übergeordnete Ziel wurde entweder als kontinuierliches Ziel mit verschiedenen Graden der Zielerreichung oder als dichotomes Ziel formuliert, das entweder nur voll und ganz oder gar nicht erreicht werden konnte. Erneut konnte derselbe Effekt des Faktors Zieltyp des übergeordneten Ziels auf die Bestrafung der devianten Gruppe wie in den vorherherigen Studien gefunden werden. Probanden in der Minimalzielbedingung waren starker geneigt die Mitglieder der devianten Gruppe monetär zu bestrafen und sozial auszuschließen. Dieser Effekt wurde erneut durch negative moralische Emotionen mediiert. Eine fünfte Studie benutzte dieselbe Zieltypmanipulation wie in Studie 4, basierend auf einem einfachen Framing des Zieltyps des übergeordneten Ziels. In dieser Studie konnte der Effekt des Zieltyps des übergeordneten Ziels auf den generalisierten sozialen Ausschluß unbeteiligter Mitglieder der devianten Gruppe repliziert werden. Eine zusammenfassende Analyse zeigt die hohe Stabilität des Zieltyps des übergeordneten Ziels auf die Bestrafung von Mitgliedern einer abweichenden Gruppe (*Zieltyphypothese)*.

Zusammengefasst trägt diese Dissertation zu einem verbesserten Verständnis der Prozesse bei, warum Mitglieder abweichender Gruppen für dieselbe Abweichung mitunter nur milde bestraft werden und sich manchmal als Reakation mit schwerwiegenden Strafen und Gewalt konfrontiert sehen. Ein ambitioniertes experimentelles Paradigma wurde entwickelt, das die Untersuchung von Determinanten von explizit negativem Verhalten gegenüber Mitgliedern einer devianten Gruppe ermöglicht. Als besonderes Merkmal können im Rahmen dieses Paradigmas verschiedene Formen explizit negativen Verhaltens gemessen und zugleich eine experimentelle Kontrolle gewährleistet werden. In diesem experimentellen Paradigma wurden zwei Hypothesen über fünf Studien hinweg getestet und bestätigt: Die *Zieltyp-Hypothese* und die *Negative Emotionen Hypothese*. Die Ergebnisse belegen, dass Individuen mit einer Minimalzielrepräsentation eines übergeordneten Ziels Mitglieder einer devianten Gruppe signifikant härter bestrafen und eher ausschließen als Individuen mit einer Maximalzielrepräsentation dieses übergeordneten Ziels.

Ehrenwörtliche Erklärung

Ich erkläre hiermit, dass mir die Promotionsordnung der Fakultät für Sozial- und Verhaltenswissenschaften bekannt ist. Ferner erkläre ich, dass ich die vorliegende Arbeit selbst und ohne unzulässige Hilfe Dritter angefertigt habe. Alle von mir benutzten Hilfsmittel, persönliche Mitteilungen und Quellen sind in der Arbeit angegeben. Bei der Durchführung der empirischen Studien haben mir folgende Personen in der jeweils beschriebenen Weise geholfen:

1. Bei den empirischen Studien I, II und III haben Tina Treml und Conni Just als studentische Hilfskräfte bei der Rekrutierung von Versuchspersonen geholfen.
2. Bei der Studie IV haben Conni Just und Nils Kupfer als studentische Hilfskräfte bei der Rekrutierung von Versuchspersonen und der Erhebung der Daten geholfen.
3. Bei der Studie V haben Nils Kupfer und Maren Patzwahl als studentische Hilfskräfte bei der Rekrutierung von Versuchspersonen und der Erhebung der Daten geholfen.

Weitere Personen waren an der inhaltlich-materiellen Erstellung der Arbeit nicht beteiligt. Insbesondere habe ich hierfür nicht die Hilfe eines Promotionsberaters in Anspruch genommen und Dritte haben weder unmittelbar noch mittelbar geldwerte Leistungen von mir für Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen. Die Arbeit wurde weder im In- noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt. Weder früher noch gegenwärtig habe ich an einer anderen Hochschule eine Dissertation eingereicht. Ich versichere, dass ich nach bestem Wissen die reine Wahrheit gesagt und nichts verschwiegen habe.

_____

Ort, Datum, Unterschrift

Bastian Lücke

* 22.05.1979

DFG Forschergruppe “Diskriminierung und Toleranz”

Universität Jena

07743 Jena

Wildstrasse 1

bastian.luecke@uni-jena.de

Akademischer Werdegang

Universität Jena, Doktorand, Betreuer: Prof T. Kessler & Prof: A. Mummendey

Universiäty of Göttingen, Diplom Sozialwissenschaften

WS 99/00 -	Hauptfächer: Sozialpsychologie, Soziologie, Betriebliche
SS 05	Finanzwirtschaft, Handels- und Gesellschaftsrecht

Universität Uppsala, Schweden

August 00 - August 01	Auslandsstudium, Lehrstuhl für Ökonomik, Erasmus-Stipendium

Praktika und Arbeiten als Studentische Hilfskraft

October 01 - April 05	Studentische Hilfskraft, University of Göttingen, Lehrstuhl für Sozialpsychologie, Schwerpunkt: Kleingruppenforschung.
October 01 - October 05	Studentische Hilfskraft am Soziologischen Forschungszentrum (SOFI), Universität Göttingen. Qualitative Evaluation eines neuen gruppenbasierten Produktionskonzepts (Auto 5000) der Volkswagen AG
August - October 04	Praktikum bei der Deutsch-Schwedischen Handelskammer, Stockholm

Arbeitsverhältnisse

2006 – aktuell	Wissenschaftlicher Mitarbeiter im Projekt: „Psychologische Grenzen V on Toleranz: Voraussetzungen und Determinanten expliziter Ablehnung und Ausgrenzung von Fremdgruppen" zusammen mit Amélie Mummendey, Thomas Kessler und Anne Berthold
April 2009 – September 2010	Wissenschaftlicher Mitarbeiter am Lehrstuhl für Sozialpsychologie, Universität of Jena
Juli 2008 – 2009	Research fellow, International Max Planck Research School on Adapting Behavior in a Fundamentally Uncertain World, Jena
Oktober 2008 - März 2009	Tutor Universität Hagen

…Thank You!

First of all I want to thank my parents who have always been there for me. Always.

I also would like to thank my supervisors Thomas Kessler and Amélie Mummendey for the inspiring discussions, the good cooperation and their friendship. I have learned a lot.

Last but not least I want to thank the members of the DFG-Research Unit "Discrimination and Tolerance" and the staff at the chair for social psychology at the University of Jena, in particular Anne, Gabi, Gerhardt, Jessica, Katharina, Lisa, Michael, Nils, Rina and Tamara, whom I consider only secondarily as colleagues, and most of all as friends.